Fall 12-16-2023

Stable complexity: verbal inflection in prominent and frequent environments

Lukas Denk

Lukas Denk
Candidate

Linguistics
Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Caroline Smith, PhD , Chairperson

Kilu von Prince, PhD

Melvatha Chee, PhD

William Croft, PhD

Melissa Axelrod, PhD

STABLE COMPLEXITY:
VERBAL INFLECTION IN PROMINENT AND FREQUENT ENVIRONMENTS

by

LUKAS DENK

B.A., General and Comparative Linguistics, Classical Philology, Media Studies, University of Regensburg, 2013
M.A., General and Comparative Linguistics
University of Regensburg, 2015

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

Linguistics

The University of New Mexico
Albuquerque, New Mexico

December, 2023

Acknowledgements

It becomes challenging to fully comprehend the various elements that facilitated the completion of this dissertation once the manuscript is written. Even for myself, it is difficult to untangle the ideas that emerged from multiple pathways and to recount the numerous instances that led me to stop and restructure the process. However, what remains is the memory of the individuals and institutions that played a pivotal role in supporting and facilitating the completion of my dissertation. As one of my advisors remarked, “You bit off more than you could chew, but you managed to chew it all.” The reason is that everyone’s input was too valuable to be left ‘unchewed’.

When I joined the linguistics department at UNM, my aspiration was to become a usage-based typologist, and as it turns out, I couldn’t have received a better education. Throughout the years, Bill Croft proved to be an exceptional mentor, nurturing my ideas and providing invaluable guidance, even when my concepts were initially fuzzy. His dedication to making my theories comprehensible and open to discussion added a depth to this thesis that extended beyond mere observations, delving into broader implications for functional linguistics. I am also deeply appreciative of Caroline Smith’s trust and encouragement in tackling the phonetic and phonological aspects of this thesis, areas I initially felt uncertain about. Additionally, having Kilu von Prince as an external member on my committee was greatly beneficial, as her insights helped structure my hypotheses, and flesh out the concepts with regards to morphology and complexity. I extend my gratitude to Melvatha Chee for supporting my research on Navajo, and whose work significantly influenced my ideas behind the question about what influences the distribution of complex morphology. Finally, I wish to express my heartfelt appreciation to Melissa Axelrod for her unwavering support since my days as a graduate student at UNM. Her incredibly cheerful and sharp mind have broadened my understanding on functional linguistics and Native American languages.

This dissertation would not have been possible without the generous financial support I received. I extend my wholehearted thanks to the Linguistics department for providing funding through the Greenberg Fellowship, a research opportunity under the National Science Foundation's Uniform Meaning Representation Project, and various teaching assistantships. I thank the Navajo program in particular for awarding me the Robert W. Young Scholarship and giving me the opportunity to teach Navajo Linguistics. The Remote Teaching Fellowship that I was awarded by the Center for Teaching Excellence also significantly improved my financial circumstances. I am also grateful to the Spanish department, especially to Eva Rodríguez González for boosting my confidence as a Spanish instructor and having me collaborate on a research project to enhance learning assessments. The Center for Regional Studies' invaluable support for my efforts in constructing a Corpus for Navajo Narratives were instrumental in funding this dissertation as well. Finally, I am grateful to Graduate Studies for the generous assistance provided through the Graduate Student Success Grant, which was essential to finalizing this thesis.

I would also like to acknowledge the support of other individuals that contributed to the concretization of the conceptual framework of this thesis. Joan Bybee's assistance in accessing the GRAMCATS databank was crucial in finding languages and verbal morphemes, and this database strengthened my idea to consider conditioning of inflection as a variable for morphological complexity. My dear friend Cat Butz who I met since the early years in Regensburg provided not only personal but also intellectual support on linguistic topics on typology, morphology, and complexity. Thankus erna whayren. Additionally, informal conversations with friends outside the field of linguistics offered invaluable conceptual support. Thought-provoking discussions on the philosophical connections between linguistic structure, historical materialism and cybernetics, especially with Lena Guidi, Josué Aciego, and Idris Robinson, were particularly enlightening. I also have to give a shoutout to my longtime friend Sophie Fromm, who not only provides a large fraction of happiness to my life, but also the capacity to formulate the chaos of my thoughts in logical steps. She helped me identify the phases in feedback cycles revolving around speakers and structure. João Neisinger's encouragement played a crucial role in motivating me to write the chapters, and his hard drive that is bullet proof to any disaster provided a safe repository for my work. I would also like to express my gratitude to my mother, Alejandra Guglielmi, for her significant contribution to my professional journey. Our yearlong discussions on linguistic topics and the disparities between my functionalist and her generativist education were enlightening, and it meant a lot for me that she read the entire draft and visited me from Germany to be at my defense.

I also owe gratitude to my family for their unwavering support and belief in me throughout my life before and during academia. My father, Ulrich Denk, has consistently shown loving support and pride in my linguistic interests from the very beginning. I am also grateful to my extended family in Argentina, Germany and Italy for proudly supporting my dream to become a linguist. Finally, I could not have undertaken the challenging task of writing a dissertation without my partner, Kyle White. Our relationship has been a wonderful journey that has provided the loving foundation to look forward as a family. I would also like to acknowledge Kyle's family from the Navajo Nation for helping us out in busy periods, including taking care of the house, kids and cats.

Stable complexity:
Verbal inflection in prominent and frequent environments

by

Lukas Denk

B.A., General and Comparative Linguistics, Classical Philology, Media Studies
University of Regensburg, 2013
M.A., General and Comparative Linguistics
University of Regensburg, 2015
Ph.D., Linguistics, University of New Mexico, 2023

ABSTRACT

Despite presenting challenges for speakers, complex linguistic features such as lexically conditioned inflection (LCI) persist across different languages. LCI forms part of not entirely predictable paradigms which require lexeme-specific knowledge to master. Moreover, LCI remains one of the oldest morphological phenomena in certain languages. Previous research has linked the persistence of such complexity to language-external factors like geographic and social circumstances of speech communities.

This dissertation delves into the question whether language-internal properties are associated with the distribution of inflectional complexity. LCI is compared with other inflectional paradigms across 41 genetically and geographically distant languages. The study shows that LCI is mostly found in phonologically prominent syllables and obligatory paradigms, suggesting that its persistence is attributable to the interaction of different levels of language structure. These findings underscore the relevance for usage-based theories to integrate structural effects into the factors that stabilize morphological complexity.

TABLE OF CONTENTS

List of Abbreviations

A	Aspect (Appendix)
A	Transitive subject/Agent
ADV	Adverbial affixes (Appendix)
AN	Animate
AUX	Auxiliary
B	Gender 'B'
BR	Basic root
C	Clusivity (Appendix)
CL	Classifier (Appendix)
Cond.	conditioning (Appendix)
DO	Direct object
DPT	Distant Past
du	Dual
E	Evidentiality (Appendix)
EPV	Epenthetic vowel
ERG	Ergative
exc	Exclusive
F	Feminine
FORM	Formative
FUT	Future
G	Gender (Appendix)
G	Grammatically conditioned (Appendix)
GCI	Grammatically conditioned inflection
GEN	Genitive
HAB	Habitual
HO	Honorific (Appendix)
HRT	Hortative
HUM	Human
IMP	Imperative
IMPF	Imperfective
INAN	Inanimate
INC	Inceptive
inc	Inclusive
INCP	Incompletive
IND	Indicative
INT	Intention
IRR	Irrealis
IS	Indefinite subject
J	Gender 'J'
L	Lexically conditioned (Appendix)
LCI	Lexically conditioned inflection
LEX	Lexical morphs (main and compound stem) (Appendix)
M	Masculine
M	Mood (Appendix)
MM	Middle marker
MOD	Modal
N	Neuter
N	Number (Appendix)
NA	Non-animate
NARR	Narrative
NDEP	Non-dependent
NEG	Negation
NHUM	Non-human
CL1	Class 1
CMP	Completive
CONT	Continuative
D	Gender 'D'
DEF	Definite
DEM	Demonstrative
DEP	Dependent
DIR	Directional
DIST	Distant
NI	Non-inflectional
NONDEC	Non-declarative
NP	Noun phrase (Appendix)
O	Transitive object
OBJ	Object
Obl.	obligatoriness (Appendix)
OPT	Optative
P	Patient (as a feature) /Position (with following number)
P	Person (Appendix)
PCI	Phonologically-conditioned inflection
PERF	Perfective
pl	Plural
PO	Polarity (Appendix)
pos.	'possible' (Appendix)
Pos.	position (Appendix)
PR	Present
PROG	Progressive
Prom.	prominence (Appendix)
PT	Past
PT2	Past 2
RECP	Reciprocal
RECPT	Recent past
S	Intransitive subject
SBJ	Subjunctive
setB	Set B indexation markers
sg	Singular
SIMUL	Simultaneous
SS	same subject
STAT	Stative
SUBJ	Subject
T	Tense (Appendix)
TAMI(P)	Tense/Mood/Aspect/Indexation/(Polarity)
TERM	Terminative case
TH	Thematic
TR	Transitive
U	Unconditioned (Appendix)
UCI	Unconditioned inflection
UCI	Unconditioned or phonologically conditioned inflection
V	Gender 'V'
V	Valency (Appendix)
VP	Vocalic prefix
X	Non-inflectional position (Appendix)

CHAPTER I: COMPLEXITY, STRUCTURAL EFFECTS, AND TYPOLOGY

1.1 Introduction

Learning a language can be hard, and this might be due to specific patterns that serve no immediate purpose for communication. Inflectional paradigms can be especially difficult to master when they contain many features and opaque forms. Whoever has taken a language course in Spanish, German or Russian has come across tables like the following one:

	traer		*venir*	
	present	simple past	present	simple past
1sg	*traig-o*	*traj-e*	*veng-o*	*vin-e*
2sg	*tra-es*	*traj-iste*	*vien-es*	*vin-iste*
3sg	*tra-e*	*traj-o*	*vien-e*	*vin-o*
1pl	*tra-emos*	*traj-imos*	*ven-imos*	*vin-imos*
2pl	*tra-éis*	*traj-isteis*	*ven-ís*	*vin-isteis*
3pl	*tra-en*	*traj-eron*	*vien-en*	*vin-ieron*

Table 1.1: Spanish conjugation of *traer* 'bring' and *venir* 'come'

At the beginning, any table presenting forms of a paradigm might appear intimidating to students attempting to master each form. However, the mere number of forms and categories is usually not a problem that lingers around for long. On a first glance, some suffixes in Table 1.1, such as *-o* '1st person singular', can be generalized and applied to further lexemes: *veng-o, traig-o, com-o, hag-o*. On a closer look, there are some intriguing differences in the stem and suffixes between *traer* and *venir*. Instructors would tell students that *traer* belongs to the 'er-conjugation', and *venir* of the 'ir-conjugation' that determines part of the forms.

Nevertheless, the criteria for assigning a lexeme to a specific class are neither phonologically, semantically nor syntactically fully predictable, which means that the assignment must be learned lexeme by lexeme. In addition to the suffixes, we see allomorphy in the stems of *traer* and *venir.* The stem paradigm of *traer* comprises the forms *tra, traj,* and *traig*, and the one of *venir* the forms *ven, veng, vien,* and *vin*. Grammars of Spanish do not usually group these alternations into different classes, and instead list them as lexical peculiarities of verbs. Ultimately, the forms of both the stem and suffixes in the paradigms listed above are conditioned by the *lexicon*, and this conditioning increases the difficulty when choosing the correct allomorph for a specific grammatical context.

Thus, Spanish verbs are especially challenging because inflection is not fully generalizable across verbs, and the pertinence to a certain pattern must be learned with the verb. Moreover, lexically conditioned paradigms are not a marginal phenomenon, they exist in other Indo-European languages, such as in English, where there are several 'strong' verbs with lexicalized inflection, such as *sing – sang – sung* and *keep – kept – kept*. And yet, this is not only a European phenomenon. Prominent examples can be found in languages all around the globe, such as in Navajo. Compare the stems *dlą́* of the verb 'drink', *kę́ę́s* of the verb 'fall across' and *áá́h* of the verb 'turn around' in Table 1.2.

Imperfective	Iterative	Perfective	Future	Optative
*yish-**dlą́*** I drink	*násh-**dlį́į́h*** I usually drink	*yish-**dlą́ą́’*** I drank	*deesh-**dlį́į́ł*** I will drink	*wosh-**dlą́ą́’*** I may drink
*nii-**kęęs*** It falls across	*niná-**kǫs*** It usually falls across	*niní-**kę́ę́z*** It fell across	*ndoo-**kǫs*** It will fall across	*noo-**kęęs*** It may fall across
*náhásh-**ááh*** I turn around	*nínáháshd-**ááh*** I usually turn around	*náhá-**yá*** I turned around	*náhideesh-**ááł*** I will turn around	*náhósh-**a’*** I may turn around

Table 1.2: Stem alternation for tense, mood and aspect in Navajo.

While there might be some shared patterns for the alternation (Eddington and Lachler 2006), such as a high tone perfective stem, each pattern is slightly different from verb to verb. Because lexicalized paradigms are not a rarity across languages (see Bickel and Nichols’ 2007 examples pertaining to ‘flexivity’), one could ask whether they serve a specific function that justifies their existence; that is, do they provide a benefit for communication that is absent in non-lexically conditioned paradigms? Many linguists doubt that there is such a benefit, as can be seen in the following statement:

> “Lexical idiosyncrasy does seem to be an irrational and counter-productive property of language. The changes that introduce it make grammar less predictable and less ordered. Lexical idiosyncrasy would thus seem to be nothing but “historical baggage” or “diachronic junk”. It may be argued, however, that lexical idiosyncrasy introduces potentially useful redundancy and therefore serves a certain synchronic function.” Dahl (2004: 112).

Further in this regard, Anderson (2015) concludes that allomorphs that are conditioned by lexical classes are more complex than the ones that are phonologically, semantically or grammatically conditioned:

> "Since phonological conditioning factors are, at least in principle, transparent, they contribute less complexity (again, in principle) than cases in which unpredictable allomorphy is based on specific morphological categories or on semantically or grammatically coherent sets of categories. These, in turn, appear less complex than ones in which the allomorphy is conditioned by (synchronically) arbitrary subsets of the lexicon, such as the Celtic mutations" (Anderson 2015: 23).

The existence of lexicalized paradigms can be approached from the angle of 'linguistic complexity' – morphology comprises more and less complex patterns, and lexically conditioned allomorphy represents a very complex type. But how do difficulty and complexity relate to one another? As will be elaborated in Section 1.2, structural complexity can be seen as one of the causes for speaker difficulty. Whereas operationalization of difficulty is useful to classify the psycholinguistic effects on speakers reacting to structure, complexity is useful to classify the structures that cause difficulty (Palotti 2019). As a result, the notion of complexity enables the creation of variables to facilitate cross-linguistic comparisons of structural types that lead to difficulty. The last decades have brought forward different ways to measure morphological complexity, with the goal of cross-linguistic comparison (cf. Dahl 2004; Baerman, Brown & Corbett 2015; Arkadiev and Gardani 2020). Whether difficult or complex – lexicalized inflection persists in languages, and this persistence is at least intriguing, especially from a linguistic perspective that is functionalist and usage-based.

This dissertation undertakes the task to research why complex patterns persist in languages despite causing difficulty and chooses to research this question by focusing on the phenomenon of lexically conditioned inflection (LCI) in verbal constructions. One would expect that paradigms that are difficult to master would eventually vanish from the language system. Against this expectation, lexicalized inflection has been frequently identified as one of the most archaic traits of language, pertaining to so-called 'mature phenomena' (Dahl 2004). In recent decades, factors contributing to difficult-to-learn, complex patterns have been attributed to extra-linguistic circumstances such as low degree of external contact and small size of a speech community (Trudgill 2011, inter alia). There is no intention for speakers to retain these complex structures, but they do survive because of favorable environments. As such, complex structures like LCI can be seen as ancient, mature phenomena (Dahl 2004). However, to a lesser extent have the factors *internal* to language been studied that contribute to the perpetuation of morphological complexity. Dahl (2004) provides a tentative explanation for why Semitic ablaut patterns, a case of complex, mature morphology, has remained unaltered for millennia:

> "Ablaut patterns have some specific properties that may contribute to their astonishing stability. Although ablaut is without any doubt a non-linear phenomenon, it is realized segmentally rather than prosodically; moreover, it is realized in a very salient way, as an alternation of stem vowels, which tend to have full stress, and exploiting basic distinctions found in almost all vowel systems. It is plausible that this makes ablaut patterns less likely to be subject to reduction processes, and it may also make them less sensitive to suboptimal transmission effects" (Dahl 2004: 274)

However, the role of structural effects in the survival of complex patterns is not further elaborated, at least not with a typological aspiration. Nevertheless Dahl's (2004) second quote above should show that structural units like LCI do not exist isolated from other structural levels of language. Likewise, the interaction between two structural units does not occur in a vacuum; speakers are always involved. The structural properties such as 'stress' that prolong the lifetime of morphological patterns must create an effect in speakers; as Dahl (2004) implies, stress achieves this by being *salient* leading to better *transmission effects* in communication.

The relationship between structure and speakers can be understood by so-called "functional-adaptive constraints" that impact the change of language structure towards conditions that facilitate communication and processing (see Haspelmath 2019: 6). These functional-adaptive constraints include preference for transparent form-meaning mappings, maximally distinctive phonemes or efficient encoding. However, if functional-adaptive constraints were to act on every level of language at the same time, one would not expect structures like LCI to arise, and less to be maintained. Lexically conditioned inflection is, among other qualities that make it less functionally-adaptive, not as transparent in comparison to other types of inflection, such as paradigms that do not exhibit allomorphy. Thus, the impact of functional-adaptive constraints must be limited or be of indirect nature. On the other hand, stability of a linguistic phenomenon leads to the realization that either the phenomenon or the environment where it thrives is beneficial for speakers; thus, the relationship between functional-adaptive constraints, structural environments and the stability of complex patterns is a study worth focusing on, and this dissertation is advanced to explore this relationship.

The subtitle of this dissertation "*Verbal inflection in prominent and frequent environments"* specifies the structural context under which the relationship between complexity and favorable structural environments is studied. I have chosen verbal inflection as the variable for inflectional complexity because verbs usually exhibit many inflectional categories across languages[1], such as tense, aspect, mood, indexation. The environments favoring inflectional complexity investigated in this dissertation are *prominence* and *frequency*. As is argued subsequently, prominence of any level in language (semantic, morphosyntactic, phonological) facilitates communication in multiple ways (easier acquisition, memorization, production). High token and type frequency also plays a facilitative role. However, prominence and high frequency do not always have the same effect on the stability of structure. Frequency can lead to longer retention due to stronger entrenchment (the degree to which a linguistic unit is firmly established in a speaker's mental lexicon) but can also lead to faster loss due to habituation (structure that is repeated too often is replaced with more effective means of transmitting information). Thus, what needs to be investigated is which type of prominence and frequency (e.g., type or token frequency) pertaining to which structural level of language (phonology, morphology, semantics) and which specific unit (phonemes, syllables, morphs, words etc.) causes structural stability. Because stable structural environments have a lasting effect on the distribution of linguistic phenomena, identifying these specific structural units will illuminate inter-structural relationships and in which structural environments morphological complexity may more likely be found across languages.

[1] According to Bickel and Nichols (2013b), the most common number of categories per verbs across languages is between 4 and 8. While there is no comparable survey regarding the number of inflectional categories for nouns or other parts of speech, it is assumed that verbs are generally morphologically richer than nouns, although there are individual languages or families where this is not the case.

This dissertation aims to accomplish the following: it tries to answer whether the distribution of inflectional complexity is influenced by structural environments associated with properties that facilitate communication. In order to investigate this question, a parameter with dependent variables for inflectional complexity and a parameter with independent variables for stable structural environments is established. To construct a parameter for inflectional complexity, different measures are surveyed that can be operationalized for variables that reflect more and less complex paradigms. This survey concludes with a parameter that classifies inflectional complexity according to phonological, grammatical and lexical conditioning of morphs and paradigms, which is used as the dependent variable.

The independent variables are prominent and frequent environments. A review of the literature to establish which prominent and frequent environments have been shown to stabilize complex morphology over time indicates that both prominent (stressed) syllables and obligatory paradigms, due to their high type frequency, are important factors. Operationalizing these variables requires various considerations that allow for the mapping between complex, prominent and obligatory morphological positions.

The relationship between inflectional complexity on the one hand and prominence and obligatoriness on the other hand is investigated in a description-based cross-linguistic study that examines the verbal morphology of 41 genetically and geographically distant languages. Promising complexity and stabilizing variables are further defined in order to map them to one another; i.e., the morphological units of verbs are aligned with different prominence and frequency values pertaining to these units. Chapter V shows that the results of most of the analyses conducted to investigate the association between more complex inflectional morphs on the one hand and prominent and obligatory morphological positions on the other hand are statistically significant. These results motivate the formulation of a theory of stabilizing structure in Chapter VI. This theory is situated within an evolutionary framework and interprets the relationship between functional-adaptive constraints and structural effects as integrated within the processes of speaker-environment interaction, utterance selection, and structure replication.

While the title summarizes the research question and can even be read as the conclusion of this thesis (complex inflection is distributed in prominent and frequent environments), the path leading to this statement requires a more elaborate discussion on morphological complexity, stability and typology. The next sections will introduce the concepts that are involved in the research question whether structural environments and the stability of these environments can also account for the stability of complex morphology.

1.2 Difficulty and complexity

As introduced in Section 1.1, this dissertation aims to answer why some patterns, which are said to be difficult to master, persist in languages. The first step in approaching this question is to narrow down what difficulty means, and how to assess it for linguistic structures. According to Pallotti (2019), difficulty is "one of the aspects that makes a task more or less challenging". Difficulty is often employed as a concept in psycholinguistic tasks (Elder et al., 2002; Fulcher & Márquez Reiter, 2003). Difficulty of a task can depend on certain linguistic features. More difficult features are the ones that "systematically appear later in interlanguage development" (Pallotti 2019: 58) or where "processing and learning requires more time and/or more mental activity" (Housen & Simoens, 2016: 166; cited in Pallotti 2019: 58). These latencies can be reflected in longer latencies to complete linguistic tasks in experiments. Thus, difficulty can pertain to structural features that exhibit difficulty, such as "a varied lexicon with low-frequency words" (lexicon) "a wide range of morphological processes" (morphology), or "a wide array of constructions with several constraints on their occurrence" (syntax).

However, there are certain reasons why difficulty is hard to operationalize for cross-linguistic research. First, difficulty is always an effect caused by speakers, and more 'difficult' language structures may not be difficult when they are presented in different tasks. Difficulty is a multidimensional construct (Pallotti 2019: 66) and is rather generalized based on tasks or speaker types than on structural types. On the other hand, the concept of linguistic *complexity* can be understood as a structure-internal property that *causes* difficulty in speakers. Pallotti (2019: 59) shows how linguistic difficulty (features that cause difficulty for speakers) can depend on linguistic complexity (related to the number and relations of structural properties), and how task complexity and task difficulty are related to these concepts (Figure 1.1).

	Complexity	Difficulty
Language	Linguistic complexity	Linguistic difficulty
Task	Task complexity	Task difficulty

Figure 1.1: Complexity and difficulty in language and tasks. (Pallotti 2019: 59)

As Figure 1.1 shows, features that are difficult to learn in language can result from their inherent complexity, complexity in the system-internal perspective, such as the number and relations of an object's structural properties (cf. Pallotti 2019: 59). Furthermore, "linguistic difficulty in turn contributes to task difficulty when a task, in order to be adequately performed, requires many difficult linguistic features" (ibid.). The difficulty of the task can be also due to the complexity of the task, and "task complexity itself may also lead of the production of more complex linguistic structures and thus contribute, in a more indirect way, to task difficulty" (ibid.).

When focusing on structural properties that cause the stability of difficult-to-learn structures such as lexically conditioned inflection, it is more plausible to assess the relative complexity of these structures. Complexity allows us to target a common factor for linguistic difficulty with the benefit of analyzing the structural properties and relationships with other structures. This is especially useful for cross-linguistic studies that operate on categories that can be found in the descriptions of the world's languages, such as this dissertation. On the other hand, generalizations of difficulty are usually assessed based on psycholinguistic experiments, and these experiments overwhelmingly include languages from a small number of language families, Indo-European languages being the most frequent. In other words, assessing difficulty typologically would involve comparing the relationship between speakers, language structure and tasks. While linguistic complexity is not the only contributor to difficulty, there are different types of complexity, and some of these might influence difficulty more than others. One goal of this dissertation is to find a parameter of morphological complexity that makes the connection between complexity and difficulty evident; lexically conditioned inflection is intuitively difficult, and as such it needs a parameter that contextualizes it as a complex structure that causes difficulty. This means that complexity must be defined from a speaker-based perspective. The following sections contextualize the speaker-based perspective of complexity within the distinction between absolute and relative complexity as well as within language-based perspectives.

1.2.1 Absolute and relative complexity

Complexity has always been a topic when discussing language, and often these discussions were full of ideological presuppositions. Kilarski (2014) demonstrates that the concept of complexity has played a crucial role in debates about various topics like variations in linguistic structures across languages, assessments of languages labeled as "exotic" or "primitive," and the presumed connections between linguistic structure and cognitive, cultural, and social factors. On the other hand, linguists have always tried to implement the concept of complexity in a more objective way, providing different interpretations (e.g., Gabelentz 2016 [1891]; Zipf 1935, 1949; Fenk-Oczlon & Fenk 2008, 2014; Givón & Shibatani 2009; Nichols 2009).

For example, complexity can refer to a constitutive property of language, both in the generative sense that an innate language faculty is able to construct complex expressions (Chomsky 1980, 1986, 2007), or that the modular, neural, and the genetic ramifications of language are complex (Pinker 1995) (see Hendrikse and van Zweel 2010 for a more detailed overview of this perspective). On the other hand, usage-based linguists are also in agreement that language is a 'complex adaptive system' (Bülow, de Bot & Hilton 2017; Beckner et al. 2009; Van Geert and Verspoor 2015; Mufwene 2008, Steels 2000) that emerges through the interaction of users, the linguistic environment, and structure.

In addition, there is a second understanding of complexity – a relative one. Language has structures that require a longer description, add to the entropy of the system, or are more difficult for speakers. This type of complexity has been widely discussed and fleshed out in recent decades (McWhorter 2001; Kusters 2003; Dahl 2004; Miestamo, Sinnemäki & Karlsson 2008; Sampson et al. 2009; Baerman, Brown & Corbett 2015; Mufwene, Coupé and Pellegrino

2017; Arkadiev and Gardani 2020). These studies are concerned with objective measures for linguistic complexity, as well as determining the factors that lead to it. So far, research on complexity has addressed every level of language, such as morphology (Baerman, Brown & Corbett 2015; Arkadiev and Gardani 2020), phonology (Easterday 2017; Maddieson 2005, 2009; Gierut 2007; Pellegrino et al. 2009), syntax (Ortega 2015; Givón & Shibatani 2009; Givón 2009; Szmrecsanyi 2004), and semantics (Maton & Doran 2017; Chersoni, Blache & Lenci 2016; Gennari & Poeppel 2003; Schoenmann 1999). Furthermore, complexity research on corpora and large language models (Park et al. 2021; Gerz et al. 2018; Cotterell et al. 2018; Mielke et al. 2019) and linear discriminative learning (e.g., Heitmeier, Chuang & Baayen 2021; Baayen et al. 2019) have provided effective tools to concretize linguistic complexity as a variable.

In Section 1.2.2, I consider speaker-independent measures of complexity. The first is description length, where the 'one-meaning-one-form principle' and the 'principle of fewer distinctions' lead to shorter description length and an increase of the minimum description length is proportional to an increase in complexity. The second measure is based on entropy, where a decrease in predictability is corelated with higher entropy and thus greater complexity.

1.2.2 Language-based approaches

The question that concerns this dissertation is how to universalize speaker difficulty as an objective variable. However, according to some linguists, bringing speakers into the calculation decreases the universal application of complexity measures. For Dahl (2004: 39), linguistic complexity must adhere to a language-based, information-theoretical understanding, and be "kept apart from other notions such as 'cost' and 'difficulty', which must always be related to a user or an agent." This allows to focus primarily on counting structural units and their interrelationships.

One of these language-based measures is *description length*, which relies on the minimum size of rules for generating/describing the data ('Kolmogorov Complexity'; Solomonoff 1964; Kolmogorov 1965). The longer the minimum description of the system is, the more complex the system is. The description can increase due to deviations from two principles: the 'one-meaning-one-form principle' and the 'principle of fewer distinctions' (Miestamo 2008). A linguistic unit that expresses one semantic feature by one specific form has the shortest description.

Paradigms with allomorphs (more than one form used for the same features) require a larger description than paradigms that do not have allomorphs. On the other hand, paradigms with syncretic morphs (the same form can express two different features) do not need a larger description than paradigms without syncretic morphs; however, a form that is syncretic needs a larger description than a form that is not syncretic because one must mention more contexts in which the form appears. Allomorphs increase of the minimum description length of paradigms and therefore their complexity, whereas syncretic morphs increase the semantic description length for a given form. Therefore, the existence of syncretic morphs and allomorphs increases morphological complexity. The principle of fewer distinctions is the paradigmatic counterpart of the one-meaning-one-form principle. The minimum description length of a system increases when a grammatical category allows a larger number of sub-distinctions. Miestamo (2008: 37) mentions an example where the past tense paradigm in Jarawara (Arauan) distinguishes between eye-witnessed and non-eye-witnessed forms; these distinctions are not found in other tenses. Therefore, the past tense paradigm is more complex than other tense paradigms; and the tense paradigm in Jarawara is more complex than a tense paradigm that does not have these sub-distinctions. Description-based measures are a good tool to assess complexity from lexicon and grammars, but they might not always reflect speaker difficulty. Nevertheless, Miestamo (2008: 28) says that "description length could, at least in principle, be of some relevance in studying cost and difficulty as well; in practice this naturally depends on how well a given theory can live up to its claims of psychological reality." A case where description length and difficulty likely align is in second-language learners who learn the grammar through rules that are listed in books. However, written rules do not predict how learners will eventually memorize and apply them.

Other speaker-independent measures of complexity adhere to predictability/entropy instead of description length. Entropy is "a measure of the reliability of guessing unknown forms on the basis of known ones" (Ackerman & Malouf 2013: 436). The greater the difficulty to predict new instances in one system, the greater the entropy, and the greater degree of entropy in a system, the greater the complexity of that system. Entropy of a system is zero if the outcome can always be predicted: this can be applied to morphological paradigms. If morphs are always predictable from the feature that they express, the paradigm has zero entropy; if there is a choice between allomorphs in the paradigm, the paradigm's entropy increases. On the other hand, Ackerman & Malouf (2013) argue that this entropy-based measure of complexity may be not far away to what speakers do when they have to choose the correct element for another given element, such as the correct affix to a given stem. While a lexically conditioned paradigm might be a prime example to show low predictability of choosing the correct morpheme given any lexical stem, it is hard to provide entropy measures in a typological survey that operates on grammars. Chapter II will acknowledge that entropy can be a measure to show that lexically conditioned inflection is more complex than other types of inflection in certain contexts, but the focus should lie on psycholinguistic evidence. This psycholinguistic dimension of complexity, i.e., complexity with regards to difficulties for speakers, is introduced in Section 1.2.3.

1.2.3 Speaker-based approaches

Lastly, there is a way to describe linguistic complexity as directly related to cost/difficulty to language users (Kusters 2003, 2008; Hawkins 2004), and this interpretation should be prioritized for this study, since the psycholinguistic effects will reflect the cross-linguistic distribution of inflectional paradigms. This user-based approach attempts to highlight the strongest connections between linguistic complexity and linguistic difficulty, as shown in Figure 1.1. However, since linguistic complexity is only one cause for linguistic difficulty, one has to find a strategy to control for as many factors as possible that obfuscate this connection, while defining variables that are applicable to all languages. According to Kusters (2008: 12–13), the "generalized outsider [...] as a second language learner, as a hearer, and as someone mainly interested in the communicative instead of the more symbolic aspects of language use" is such a limitation that can show more consistent effects between specific structures and speakers. Kusters (2008) agrees with Trudgill (2001: 371) that this type of complexity is best defined as difficulty for *second language learners*:

> "[P]henomena that are relatively difficult for a second language learner in comparison with a first language learner are the most complex. Phenomena that are easy to acquire for a second language learner but difficult for a first language learner are the least complex. The notion

> ‘generalized’ prevents positive and negative interferences of an accidental first language to cloud our account of complexity.” (Kusters 2008: 9)

This definition avoids the variable manifestation of difficulty in L1 speakers present during the process of acquisition. L1 acquisition has many more stages than L2 acquisition, and as such, the difficulty to acquire certain phenomena depends on the specific stage – some phenomena can have opposing effects on L1 and L2 speakers; what is difficult for L2 speakers can provide a benefit for L1 learners. An example of this is redundancy. Morphology that repeats features can be an additional burden for L2 learners in the sense that they must learn when the repetition is used, whereas L1 learners primarily benefit from this repetition (Trudgill 1999; Wit & Gilette 1999; McWhorter 2001, Sagarra 2001; Lupyan and Dale 2010). Once a ‘generalized outsider’ is defined, the connection between linguistic complexity and difficulty is more consistent, and this allows to construct variables that are more robust and lend themselves to wider application.

According to Kusters (2008: 13), the generalized outsider favors structures that are maximally *economical*, *transparent* and *isomorphic*. Economy refers to the number of categories and category combinations in the language – a lower number is easier to learn and does not add much complexity to the system. Transparency refers to the form-meaning relation in morphemes – a greater number of allomorphs and homonyms poses difficulties for L2 learners. Isomorphy refers to how consistent the order of semantic and syntactic categories is across constructions – “the more the order in the morphological domain is computable and motivated by the order in the semantic or syntactic domain, the more isomorphic the morphology is” (Kusters 2008: 13). Parallels between syntax and morphology facilitate the learning for L2 speakers. All these parameters may not cause difficulties in L1 acquisition. Kusters (2008) abandons the maxim that complexity measures must be completely objective since, as he argues, objectivity cannot even be guaranteed for absolute, language-based measures, as they are dependent on the linguistic analysis and theory. Andrason (2014: 85) adopts a more radical view, in which he argues that an objective measure for linguistic complexity is impossible to determine since languages are already "infinitely complex". However, by setting the vantage point in second language speakers, speaker-based complexity might be measured in the most objective way and capture what drives the interest in linguistic complexity in this thesis, namely speaker difficulty.

Nevertheless, difficulty exists for L1-learners too, and there is no reason not to consider this difficulty when determining linguistic complexity. After all, acquisition of language by L1 might not be very different from immersive acquisition of language by L2, and instead be connected (Marinova-Todd, Marshall & Snow 2000; MacWhinney 1999; Van Geert 2008, inter alia). In both cases, (complex) morphology is mastered at a later stage (see Clark 2017 for acquisition of complex morphology in children). From a usage-based approach, language is shaped both by children and adults, and difficulty experienced by children impacts the reproduction and propagation of certain patterns to other members of the speech community. The 'complex-adaptive-system' view of language treats both children and adults as selectors of utterances and drivers of change.

Because this dissertation examines the distribution of morphological complexity in the languages of the world, and distributional asymmetries arise due to language usage, it will prioritize a speaker-based approach to complexity. In Chapter II, evidence from both L1 and L2 learning is considered to justify the determination of relative complexity in inflectional paradigms. Furthermore, information-theoretical arguments based on description length and entropy will also be invoked if there are reasons to connect them to speaker-based complexity.

1.3 Structural effects

In Section 1.2, it was argued that linguistic difficulty can stem from linguistic complexity – structures that cause difficulty for language users. This implies that some complex features in language evade functional explanations based on principles that facilitate communication. Functional principles aim to understand why certain constructions exist and how they serve the needs of language users. When functional principles are not immediately evident for specific constructions, diachronic explanations are preferred, attributing changes in linguistic structures to past functional influences. Some linguistic phenomena, such as interrogative constructions and phonetic assimilation, have clear functional explanations (obtaining information and facilitating articulation respectively). However, other phenomena, like the marking of participants in transitive events, require more elaborate answers that may not always be universally agreed upon. Cristofaro (2019) argues that linguistic systems such as morphosyntactic alignment between intransitive and transitive participants may not have a unifying function but are the result of diachronic developments of constructions that served different functions in the past. Conversely, Haspelmath (2019) proposes result-oriented "functional-adaptive constraints" that explain how languages develop properties that facilitate communication. By defining these constraints, factors such as acoustic space or the association between semantic roles and prominence status are considered. Lexically conditioned inflection, as presented in the introduction of this chapter, poses a challenge for these functionalist approaches. Vestigial traits, like lexically conditioned allomorph paradigms, may be too ancient to trace back through specific developmental paths. These complex traits lack evidence for present-day functional-adaptive constraints and can pose difficulties for speakers.

While functionalist approaches may explain the presence of difficult-to-learn morphology due to geographical and social factors, less is known about the distribution of this type of morphology in utterances. Chapter III aims to investigate how structural asymmetries influence its change and stability. It is expected that all stable units exhibit properties that facilitate communication, but not all facilitative properties make structure stable. Thus, in order to search for structural units that stabilize (complex) morphology, one could search for facilitative structural properties and research which of these are also associated with stability. Chapter III investigates two properties, prominence and entrenchment, and seeks to identify which prominent and entrenched structural units are more likely to preserve inflectional complexity.

Prominence refers to the contrast and effectiveness of conveying information through prosodic, morphological and semantic structure, while entrenchment relates to the degree to which a linguistic unit is firmly established in a speaker's mental lexicon. Both prominence and entrenchment have functional-adaptive aspects that facilitate communication. While one can deduce from grammatical descriptions which parts of utterances are more prominent than others (for example stressed syllables, word boundaries, stems), degree of entrenchment is not easily identifiable. However, frequency of occurrence is a factor closely associated with entrenchment; more frequent linguistic units are more entrenched. Frequency itself is not a structural factor like prominence and entrenchment, but due to the correlation between frequency and entrenchment and the existing literature on frequency effects, frequency will be chosen as a proxy to determine the effects of entrenchment on language structure. In sum, the focus of Chapter III is to explore the facilitative and stabilizing effects of different prominent and frequent units on language structure. Studies on acquisition, production, perception, and memorization provide insights into their facilitative potential, while diachronic studies shed light on their stabilizing potential. By understanding which prominent and frequent units can stabilize morphology, this research sets the stage for cross-linguistic analysis and proposes variables for further investigation.

The relationship between facilitative and stable structures can be exemplified as follows. High token frequency of a linguistic unit helps retrieve this linguistic unit more easily, but as an effect of token frequency, the unit may weaken and disappear. This erosive process of high token frequency can affect individual phonemes (such as the change from /s/ to /h/ to zero in Spanish: *los chicos* > *loh chicoh* > *lo chico*) or morphemes (*God be with you* > *good-bye*). On the other hand, high *type* frequency of words and morphemes is more likely to preserve linguistic units (such as the *-ed* suffix for past in English). With regards to prominence, stressed syllables are not only facilitative in the sense that children acquire them earlier than unstressed syllables, but stress can also cause syllables to be stable. This can be seen in the development of ‘vuestra merced’ to ‘usted’ in Spanish (De Jonge and Nieuwenhuijsen 2009: 1651, quoted by Sáez Rivera 2013: 114):

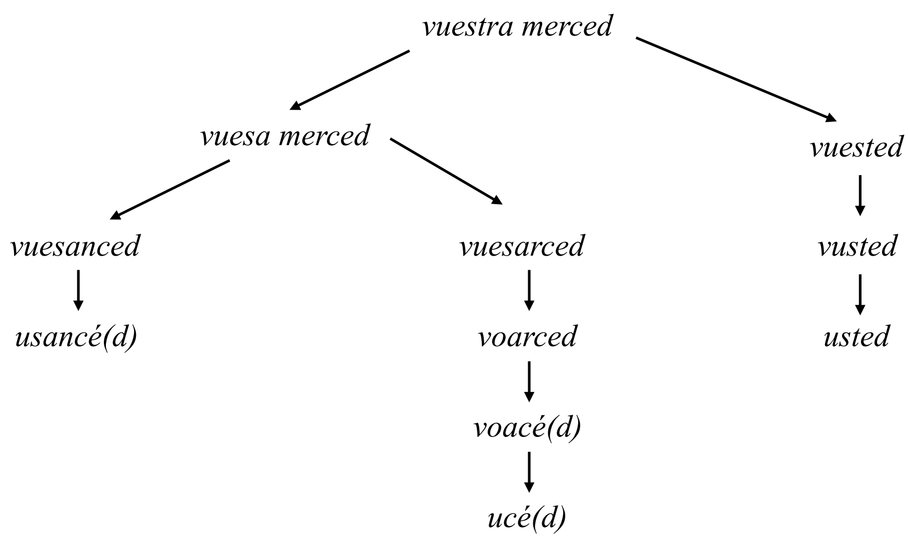


Figure 1.2: Evolution of the phrase *vuestra merced* 'Her Mercy' to '*Usted*' '2nd person singular formal' (Sáez Rivera 2013: 114).

The different stages show the likelihood for syllables to be preserved in *vuestra merced*. The syllable *us* [us] is retained as it originates from the stressed syllable *vues* [vwes], and *ed* [eð] is also retained from the stressed syllable *sed* [seð]; the other segments belonging to unstressed syllables have mostly disappeared or shifted. Thus, stressed syllables are, unlike the unstressed syllables, a stable structural environment. Chapter III will argue that prominence and frequency/entrenchment alone does not explain the relative stability of a unit (and therefore its asymmetric distribution in the languages of the world). Besides functional-adaptive constraints, the effects of the structural unit itself that is prominent and entrenched needs to be taken into consideration in order to determine the environments where morphological complexity survives. The relationship between functional-adaptive constraints and structural factors is further interpreted in Chapter VI from an evolutionary perspective. However, the most challenging task in this dissertation consists of translating prominent and entrenched/frequent environments into variables that can be operationalized for a cross-linguistic comparison. The following section introduces the typological approach chosen in this thesis to illuminate the relationship between inflectional complexity and stable structural environments.

1.4 Typology

This study undertakes a cross-linguistic, description-based, typological investigation as a way to define the relationship between inflectional complexity and structural environments that stabilize said complexity. Typological approaches are useful to show universal patterns in languages, which "call for explanation in terms of more general cognitive, social-interactional, processing, perceptual or other abilities" (Croft 2003: 5). While typological approaches are principally agnostic to a specific theory (Nichols 2007), most of them follow functional assumptions since functions have proven useful to explain structural diversity. The explanations are based on "competing motivations, economy, iconicity, processing, semantic maps in conceptual space" (Croft 2003: 3). Typological studies in languages are bottom-up (empiricist/inductive) approaches where different phenomena are compared to one another and universal patterns are identified (Bloomfield 1933; Aikhenvald and Dixon 2007; Croft 2003; Ramat 2011; Moravcsik 2012). These patterns can be formulated through implicational universals or hierarchies (Greenberg 1966; Corbett 2000; Croft 2003). This contrasts with top-down approaches that presuppose underlying universal structural categories that instantiate surface phenomena, such as in generative grammar (e.g., Newmeyer 2007). While the field of typology has been growing in the last decades, it still lacks precise definitions of concepts that are universally applicable. According to Haspelmath (2010: 664), typology cannot formulate universal categories, but must posit 'comparative concepts' that are "specifically designed for the purpose of comparison [and] (…) are independent of descriptive categories." These comparative concepts refer to semantic categories such as 'agent', 'patient'; 'object', 'event', or general formal categories such as 'precede', 'overt' 'identical' (Haspelmath 2010: 670). Thus, the tendency in typology is to explain form-meaning units (morphology, syntax) by attributing general semantic and phonological principles, not explaining form or meaning based on universal morphological and syntactic categories. However, since this thesis seeks to find generalizations of the interplay between morphology and structural stability, one cannot apply only semantic or phonological categories. Instead, it must find categories where the specific form-meaning relationship can be compared.

1.4.1 From comparative concepts to variables

This thesis faces the challenge that morphological elements need to be generalized to be used as variables. Since I choose to apply a usage-based functional approach, I will avoid positing universal syntactic and morphological categories. The usage-based approach is *emergentist* in nature, and views morphological and syntactic categories as derivative from more general principles, among them phonological and semantic matches (e.g., Bybee 1985; Langacker 2008, 2019). Nevertheless, one can also generalize the *effects* that certain form-meaning units have on speakers, or how the interaction affects the change of these units. Structure can be classified and operationalized according to consistent behaviors they evoke in interlocutors. However, this bears another challenge: a cross-linguistic study must use reported effects from other studies to construct 'shortcut generalizations'. For the purpose of this dissertation, parts of utterances need to be classified according to user effects and stabilizing potential, and grammars usually do not mention these properties. Thus, what is needed for the cross-linguistic comparison is a deductive step in which reported effects of inflectional paradigms, prominence and entrenchment in some languages are generalized for the purpose of creating psycholinguistically based comparative concepts to construct variables for a description-based typology. Figure 1.3 shows how inductive and deductive processes were applied for the purpose of this study.

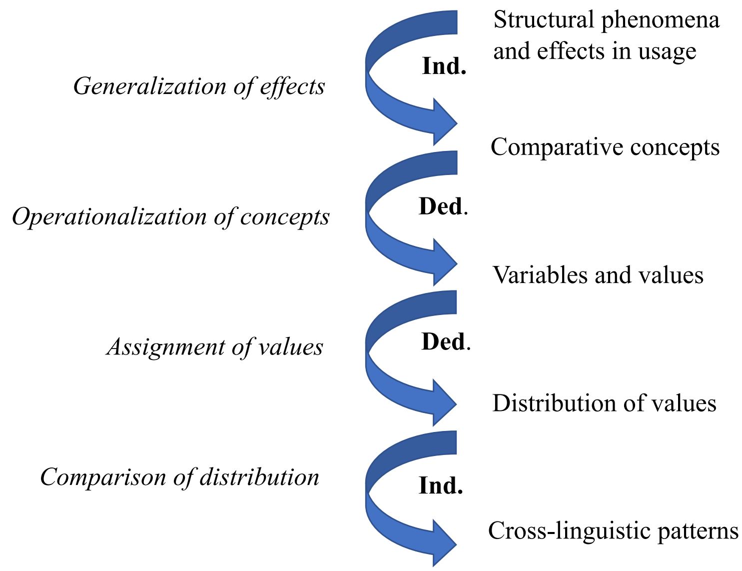


Figure 1.3: Inductive and deductive processes applied in this dissertation.

While the effects were generalized based on existing psycholinguistic and diachronic studies, the variables were constructed to facilitate collection of data from grammars. The research question of this dissertation, namely whether difficult-to-learn inflectional paradigms are distributed across prominent and frequent structural environments, is tested by an inductive method, by investigating how complex inflectional paradigms are distributed across stabilizing structural units. Therefore, although the approach used to test the hypothesis in this study is primarily inductive, the inclusion of two deductive steps in the epistemic process may diminish the validity of the results. Nevertheless, it is a necessary method aimed to uncover the connection between psycholinguistic, diachronic and comparative questions. The next section provides an overview of how the parameters ‘inflectional complexity’ and ‘stable structural environments’ have been generalized and operationalized in this dissertation.

1.4.2 Selection and assignment of values

To understand why complex paradigms like lexically conditioned inflection survive in language, it is necessary to contextualize lexically conditioned inflection within a parameter that captures other types of inflection that are considered not as difficult to learn as LCI. The complexity of LCI may stem from several properties, but only a few can be applied to construct variables that exhibit relatively consistent effects of difficulty on speakers and can be operationalized.

In Chapter II, it is argued that complexity of allomorphy, as sketched out by Anderson (2015) approximately fulfills these two criteria. Complexity in allomorphy can be attributed to different types of *conditioning* of allomorphs: phonological, grammatical, or lexical. As will be demonstrated, unconditioned and phonologically conditioned allomorphy do not pose many difficulties for speakers and hearers, since phonologically conditioned allomorphs usually facilitate the articulation and perception of utterances. On the other hand, lexically conditioned allomorphy is harder to master than non-lexically conditioned allomorphy, both for children and adults. Speakers must not only be aware of form and function of lexically conditioned allomorphs, but also of their unpredictable environment. An intermediate category in terms of difficulty is represented by grammatically conditioned allomorphy, where interlocutors must process and retrieve the correct form and meaning but also the grammatical category that conditions it. In contrast to lexical conditioning, grammatical conditioning is generalizable. Furthermore, it has been decided to focus on inflectional paradigms as they show more consistent effects across speakers. Lexically conditioned inflectional paradigms contribute particularly to complexity because users have to apply procedural and/or declarative knowledge to master the forms, and this results in hesitation and thus recorded latencies in experiments.

Nevertheless, allomorphy alone might be a too narrow concept to construct complexity degrees of inflection, and it has been decided to include other types of forms that can be also defined as conditioned by different environments. These include sub-morphemic segments that are affected by stem alternation, morphs whose features are dependent on different grammatical contexts (syncretism and distributed exponence) and cumulative morphs, which can be analyzed as a morph that incorporates conditioning and conditioned features. As such, the tripartite distinction of Anderson 'complexity of allomorphy' is expanded to complexity of conditioned inflectional paradigms in general. The complexity variables have been operationalized as unconditioned/phonologically conditioned inflection (UCI), grammatically conditioned inflection (GCI) and lexically conditioned inflection (LCI), reflecting an increase in inflectional complexity.

Since not all languages contain every inflectional type (and some of them lack inflection in general), the composition of the language sample was oriented towards languages containing the most complex type, LCI. The main sample constitutes 30 genetically and geographically distant languages with LCI. Two additional control samples were compiled: one with 6 languages exhibiting GCI and UCI but not LCI, and one with 5 languages exhibiting only UCI. In an Excel spreadsheet that is shown in the Appendix, the verbal morphology of all languages was outlined into different positions (prefixes, stem, suffixes) to which the values UCI, GCI, LCI and NI ('non-inflectional') were assigned.

Chapter IV (Methodology) discusses how these positions were defined. Since not every language encodes its inflectional categories in fixed morphological positions, different criteria were applied to determine whether one or multiple positions need to be accounted for. For example, if tense is expressed by a morph that is an enclitic in some contexts and a proclitic in others, it is discussed whether one position is the more frequent one (in which case the less frequent positions were ignored). Furthermore, Chapter IV proposes a distinction between inflectional and derivational morphology by arguing that inflectional categories are more consistent in their semantic generality and frequency of occurrence than derivational categories.

The second phase of operationalization involved identifying stable prominent and frequent units. As mentioned above, prominence facilitates the acquisition and memorization of units, but stressed syllables in particular are also more likely to persist in languages, unlike other prominent units. Because not all languages have stress, other prominent properties of syllables and segments were investigated with regards to their stabilizing potential. These were tonal contrasts and syllable weight for languages in which stress is not reported. High type frequency, which leads to deep entrenchment and stability of units, was operationalized as morphological obligatoriness. It is argued that morphs of a morphological paradigm that is obligatory will have a higher type frequency than morphs of a non-obligatory paradigm.

The verbal templates outlined for morphological positions (see Appendix) were mapped to values for prominence and obligatoriness. For languages that have stress, stress was chosen as the positive value of prominence. Positions that are always stressed received the value 'yes', positions that are sometimes stressed the value 'possible' and positions that are never stressed the value ‘no’. The same values were used to assess prominence in tone languages: ‘yes’ depicts positions that can receive any tone or the most prominent tone (subject to grammatical information on the prominence or markedness of tones), ‘possible’ was assigned to positions with morphs that have several tones but not all or the most prominent tone, and ‘no’ was assigned to positions with only one tone. For non-syllabic morpheme positions, the values depend on whether morphs can appear at word boundaries, which is an additional parameter of phonological prominence. In sum, the values ‘yes’, ‘possible’ and ‘no’ depict different degrees of phonological prominence that can be mapped with different degrees of morphological complexity.

The values used for obligatoriness capture the potential of one position occurring in every verbal position. ‘Yes’ is assigned if the position is always filled by phonological material, and ‘(yes)’ is assigned if the position is always filled but the paradigm contains zero morphs. ‘no’ appears when the position is not always filled (the grammatical category is not obligatory for all verbs). For example, a language that has subject indexation in all verbs without zero in third person is assigned the obligatoriness value ‘yes’ in this position. Because obligatoriness is an operationalization of high type frequency, suffixes that occur in marginal constructions (such as irregular verbs) were ignored.

The processes of generalizing and operationalizing effects associated with specific structures result in various decisions that are justified in Chapter II, III and, more concretely, in Chapter IV. The main hypothesis resulting from the establishment of variables (“more complex inflection is found more often in more prominent and obligatory positions”) is tested in Chapter V.

1.4.3 Evaluation

Chapter V tests the research question whether complex inflection is distributed across prominent and frequent morphological environments, by formulating two main hypotheses consisting of several sub-hypotheses. The three rows of annotation (inflectional complexity, prominence and obligatoriness) are compared and the association of the values evaluated. Chi-squared tests and one Fisher's exact test were used to determine the significance of the results. The results show a pattern that conforms to the general hypotheses: not only is more complex inflection (such as LCI and GCI) associated with more prominent positions (e.g., stressed syllables) but also with obligatory paradigms. This relationship is further researched with regards to factors that might lower the validity of results, such as root/stem[2] morphemes or the relationship between prominence and obligatoriness itself. Even after controlling for these factors, the general pattern conforms to the ideas that have been presented in Chapters I through IV. Because most of the results are statistically significant, I included a further discussion on the relevance of these findings for an evolutionary approach to language (Chapter VI).

1.4.4 Revisiting the Theory of Utterance Selection

Finally, this dissertation also provides an interpretation of the results with regards to Croft's (2000) Theory of Utterance Selection, an evolutionary perspective on how structure in utterances is selected and replicated across speech events. Because the results in Chapter V are

[2] Throughout this thesis, 'stem' and 'root' will be used synonymously; 'stem' will be the preferred denomination, because it suggests that it can be further decomposed into segments. 'Stem' does not mean root plus derivational affixes, as in some grammatical descriptions. 'Root' will be used when referencing other authors.

statistically significant, they strengthen the idea that some complex phenomena are stabilized by certain linguistic environments, rather than established due to purely functional motivations. It is rather the specific structural units involved, and their particular functional prominence, and entrenchment due to high frequency, that promote and stabilize difficult-to-learn inflection. Thus, environments favoring the propagation and stability of complex inflection are not only external to language, such as close-knit communities in residual zones (see Nichols 1992), but exist in the structure of language itself. This calls for a revision of the role of structure from a functionalist, evolutionary perspective. In particular, structure is not only a result shaped by the various forces of interaction between speakers and the environment, but it also causes reverberation in the system. According to the Theory of Utterance Selection formulated in Croft (2000), speakers interact with their environment, and this leads to differential replication of structure, where some structures are preferred over others (i.e., selected). While accommodating the structural effects obtained in the empirical part of this dissertation within the Theory of Utterance Selection, I argue that replicated structure is not only part of utterances that speakers select, but also part of the environments with which speakers interact, which in turn can influence the rate at which differential replication happens. This differential replication is reflected in the differential distribution of complex inflection in prominent and frequent environments. It is necessary to consider the structural aspect of these effects because prominence and entrenchment alone do not influence the distribution of complex features in language; in fact, the specific structural units that are prominent and frequent do.

1.5 Thesis outline

This introductory chapter has provided an overview of the theoretical and methodological insights of the doctoral dissertation. The following chapters will concretize the concepts of inflectional complexity, facilitating and stabilizing structure, and the typological approach used in the investigation.

Chapter II introduces the concept of complexity of conditioned inflection and shows to which extent this concept reflects language- and speaker-based complexity. Chapter III surveys evidence for the facilitative and stabilizing potential of prominent and highly frequent units on morphology. Chapter IV explains how the languages for the typological survey were compiled and classified into three samples. This chapter also proposes an operationalization of complexity of inflectional paradigms, prosodic prominence and high type frequency in order to be used as variables. Chapter V formulates the hypotheses related to the research question and evaluates the association between the variables. Chapter VI interprets the results from the evolutionary Theory of Utterance Selection. Finally, Chapter VII summarizes this thesis and states the implications of the results for the theoretical assumptions made throughout this study and the contributions for the field of linguistics, as well as future research avenues. Additionally, an appendix is incorporated in this dissertation, wherein the languages and their verbal templates annotated in the Excel sheet are displayed in alphabetical order.

CHAPTER II: COMPLEXITY OF CONDITIONED INFLECTION

2.1 Introduction

User-based approaches to complexity can more likely explain the differential distribution of difficult-to-learn patterns in utterances across languages, but language-based approaches may be easier to operationalize. The problem that typologists face is that language comparison aims to generalize the diversity of structure by stating general principles, but these principles, such as hierarchies or conceptual maps, are the product of usage dynamics, not necessarily the preconditions for usage. Sinnemäki (2014) depicts this relationship between cognitive processing, usage and language variation as an emergent relationship, where processing preferences influences language acquisition and language use, language use determines language change, and language change determines the typological distributions. "[T]ypological distributions come to indirectly mirror cognitive bases' (Sinnemäki 2014: Figure 1). Another way to depict the connection between psycholinguistics and cross-linguistic research is that "principles of performance can be used to make predictions for patterns of grammatical variation, while preferences in grammars become relevant for the testing of psycholinguistic ideas" (Hawkins 2007: 87). As such, typological distributions are often explained by recurring to psycholinguistic effects, and psycholinguistic studies use cross-linguistic insights to formulate their variables (Hawkins 2007; Sinnemäki 2014; Bornkessel-Schlesewsky et al. 2008; Moravcsik 2011). However, one question that has not been fully researched is how psycholinguistic preferences can be associated with certain structural units that can be generalized to formulate typological variables.

Processing-based notions and typological categories are qualitatively different from one another, and they overlap only in few contexts. For example, one cannot say whether 'agents' or 'patients' (two widely used typological concepts) are mostly preferred in language;

speakers express these categories depending on the situational context. However, this thesis argues that certain types of complex morphology, if correctly defined, can be associated with user preferences.

Chapter I has presented language- and user-based approaches to define and measure morphological complexity. This chapter proposes arguments for choosing different types of conditioning of inflectional allomorphs as a practical parameter that can be operationalized as different degrees of user-based complexity that cross-linguistically capture relatively consistent effects of speaker difficulty. The variables of this parameter must be distinct enough to allow annotation, and general enough to allow searches in as many grammars as possible. Second, they should capture a gradient relationship to complexity between one another. This allows to classify the variables as more and less complex, while remaining distinct. Third, they should reflect user-based difficulty since this is what influences the distribution of complexity in utterances across languages. Language-based approaches to complexity can still be integrated if they reflect user-based constraints.

In Section 2.2, Anderson's (2015) parameters for morphological complexity are chosen to contextualize lexically conditioned inflection, and *complexity of allomorphy* is chosen as a starting point to classify inflectional paradigms. Because the concept of *allomorphy* lacks clarity, a definition based on conditioning types is offered (Section 2.3). These types are further elaborated in Section 2.4: phonologically, grammatically and lexically conditioned allomorphs. In Section 2.5, the relationship between these types and speaker-based difficulty is assessed, taken into consideration insights from information-theoretical and psycholinguistic perspectives. The assessment yields a cline presented in Section 2.6 in which phonologically conditioned inflection adds less user-based complexity to paradigms than grammatically conditioned inflection, and grammatically conditioned inflection add less complexity than lexically conditioned allomorphs. This cline should be also reflected in the typological distribution of inflectional allomorphs. Section 2.6 concludes that that these types can be used as variables for further typological research on morphological complexity.

2.2 Contextualizing lexically conditioned inflection

In Chapter I it was claimed that lexically conditioned inflection represents a feature of language that is relatively difficult to learn for speakers. The purpose of this chapter is to search a reliable cross-linguistic parameter to contextualize LCI as more complex than other types, and to use these types as variables that can be compared with other structural properties of verbal constructions that will be introduced in Chapter III.

When talking about lexically conditioned inflection as a difficult-to-learn phenomenon, one has to talk about morphological complexity in general. The question is: is inflectional complexity only one aspect of morphological complexity in general or does it represent a more relevant case that usually evokes attention? Arkadiev and Gardani (2020: 5) point out that morphological complexity can be read as “overall contribution of morphology to complexity”; this view is shared by several researchers, such as Aronoff (1998, 2015), Carstairs-McCarthy (2010) and Anderson (2015). Morphology is the addition of ‘internal structure’ to words, and words that have less internal structure are less complex. Morphology most clearly appears as additional baggage to second language learners – which is the vantage point for measuring linguistic complexity for Kusters (2003, 2008). However, there is less evidence that morphology is relatively difficult for first language learners. Especially the redundant character of some inflectional categories could play a beneficial role in reducing uncertainties within syntax

(Moscoso del Prado Martín 2011). Whether morphology is the more complex strategy to express form-meaning relationships (as opposed to syntax) should not be answered in this study. Nevertheless, there has been increased research on morphological complexity (Arkadiev and Gardani 2020; Moscoso del Prado 2011; Baerman, Brown & Corbett 2015; Nichols 1992) which may allow us to contextualize lexically conditioned inflection as a type of complex morphology. In the following, it was decided to examine Anderson's (2015) dimensions of morphological complexity, since they provide a systematic approach that helps contextualize lexically conditioned inflection as a complex phenomenon.

2.2.1 Dimensions in Anderson (2015)

Anderson (2015: 17-24) provides a detailed language-based approach to morphological complexity for cross-linguistic classification, which may be also helpful to capture morphological difficulty for speakers. This approach proposes several parameters, subsumed under two dimensions:

1. Overall system complexity
    - 1.1. Number of elements in the system
    - 1.2. Number of affixes in a word
    - 1.3. Principles of morphological combination

2. Complexity of exponence
    - 2.1. Complexity in the realization of individual elements
    - 2.2. Complexity of inter-word relations
    - 2.3. Complexity of allomorphy

The first dimension of morphological complexity is 'overall system complexity', which includes two quantifiable parameters, *number of elements in the system* and *number of affixes in a word*. The larger the number of elements (i.e., morphemes) in the system and affixes in words, the higher the morphological complexity. The third parameter, *principles of morphological combination*, relates to morpheme order. Morpheme order does not contribute to complexity if it "follows from the content properties of the elements involved" (Anderson 2015: 19), i.e., if affixes that have scope over more affixes are placed at the periphery of the word whereas affixes that have scope over few affixes are placed next to these. Complexity increases when the order of morphemes is not predictable from these scopal relations, but instead requires further specification from other levels of language, like phonology or semantics, as is the case in Athabascan languages (Rice 2000). This parameter can be associated with entropy-based measures since it relates to combinations between morphemes and deviations from a predictable order. It entails less complexity based on description length, since templatic tables are outlined for any morpheme orders, whether they show iconicity between semantics and morpheme order or not.

The second dimension is 'complexity of exponence'. This dimension relates to the one-meaning-one-form principle where "a single discrete, indivisible unit of form linked to exactly one discrete unit of content" (Anderson 2015: 21) does not contribute to morphological complexity, unlike a deviation from this principle. A deviation implies either more formal elements for one meaning element or more meaning elements for one formal element and can be interpreted as an increase in description length. The first parameter is *complexity in the realization of individual elements*. Complexity increases by cumulative morphs such as Latin *-ō* (1st person, singular, present) in *am-ō* 'I love' (one suffix – three features) or discontinuous morphemes, such as circumfixes, infixes (in which case the stem morpheme is discontinuous) or other cases of multiple exponence, like double negatives found in several languages (more than one formal element, same features). The parameter *complexity of inter-word relations* depicts the relation between the morphosyntactic representation and the word form. *Syncretism* "describes the situation in which multiple morphosyntactic representations map to the same word form for a given lexeme", whereas with regards to *variation*, "multiple word forms correspond to the same morphosyntactic representation" (Anderson 2015: 22). An example for syncretism is the English word [hɪt], which is the present or past tense form of 'hit'; an example for variation is the alternatives [dowv] and [dajvd] for the past tense of 'dive'. Thus, both syncretism and variation add to morphological complexity as they require a longer formal or semantic description. The third parameter is *complexity of allomorphy*. Allomorphy goes against the one-meaning-one-form principle by having multiple forms assigned to one abstract morpheme.[3] This deviation can again contribute to morphological complexity in terms of description length, since the grammar must list all the allomorphs for a given morpheme. However, rather than by description length, "allomorphy can contribute to the complexity of the system [...] depending on the bases of the principles underlying its conditioning" (Anderson 2015: 23). The conditioning contexts are qualitatively distinct, since they relate to phonology, morphology, syntax and semantics. This property aligns with the requirement formulated in Section 2.1, according to which comparative variables must be distinct enough to allow annotation and detection in Grammars.

According to Anderson (2015: 23), phonological conditioning does not increase morphological complexity since "phonological conditioning factors are, at least in principle,

[3] In Section 2.3, I discuss why this definition of allomorphy is problematic and should be replaced.

transparent." Instead, allomorphy contributes to complexity when the conditioning "is based on specific morphological categories or on semantically or grammatically coherent sets of categories." This type of conditioning is, again, less complex than when "allomorphy is conditioned by (synchronically) arbitrary subsets of the lexicon" (Anderson 2015: 23). The manifestation of the latter type are irregular allomorphs or arbitrary inflectional classes which cannot be derived from phonological or semantic/grammatical features alone; this refers to lexically conditioned allomorphs as introduced in Chapter I, which Anderson assumes to represent "perhaps the summit of complexity."

2.2.2 Applicability of complexity of allomorphy

Anderson's (2015) dimensions and parameters are a useful outline to implement morphological complexity variables in typological research. That is, one could investigate languages according to the number of elements in a system, number of affixes in a word, etc., and correlate these to other structural entities as opposed to language-external variables. Since the focus of this dissertation is inflectional morphology, and specifically the effects yielded by lexically conditioned inflection in contrast to other types of inflection, one must find a way to accommodate Anderson's complexity parameters to this distinction. The reason why this inflectional context is chosen is because, as will be further explained in Section 2.5.3 of this chapter, inflectional categories provide a similar degree of semantic generality (unlike derivational and lexical morphs) and this property functions as a controlling instance to detect the effects of morphological complexity on speakers. The questions that arise are a) which parameter is useful to classify the *inflectional* complexity presented in Chapter I (lexically conditioned vs. non-lexical conditioned inflectional paradigms); b) which parameter reflects best degrees of user-based complexity; and c) which parameter can be applied to *all* the inflectional morphology of *all* languages.

At a first glance, it seems plausible to select the last parameter, *complexity of allomorphy*, to contextualize the complexity of lexically conditioned inflection, because LCI is "allomorphy [that] is conditioned by (synchronically) arbitrary subsets of the lexicon" (Anderson 2015: 23); furthermore, Anderson (2015) clearly positions this type of allomorphy as more complex than other types. Less complex allomorphs are conditioned by semantic or grammatical categories, and even less complex allomorphs by phonological rules alone. However, less than allomorphy itself – which is an instance of form-meaning-mismatch and can also be classified under "variation" from the parameter *complexity of inter-word relations* – it is the conditioning of allomorphy that makes allomorphs more and less complex. The question is whether this tripartite distinction of conditioning could also be applied to inflection in general, which would generalize the complexity to not only allomorphs, but to cumulative, syncretic, and discontinuous morphs, or even entire paradigms.

Intuitively, lexically conditioned inflection is not only difficult because it has separate allomorphs, but because it involves also learning two alternative feature sets for the same form (syncretism), lexically conditioned discontinuous exponence, and different *sets* of (allo)morphs, such as lexical conjugation classes. These other features are indeed captured in the other parameters of Anderson's (2015) dimension 'complexity of exponence', namely *complexity in the realization of individual elements* (cumulative and discontinuous morphs) and complexity *of inter-word relations* (syncretisms and variation). This means that either, one has to find another parameter that captures specifically inflectional complexity, or one has to modify Anderson's (2015) parameter *complexity of allomorphy* to include other types of conditioned inflectional morphs. Because *complexity of allomorphy* provides a clear-cut distinction between phonologically, semantically/grammatically and lexically conditioned allomorphs, it has been decided to explore this parameter with regards to speaker difficulty. Complexity of allomorphy may capture a large portion of why lexically conditioned inflection is difficult, and allomorphy is very frequent – according to Carstairs McCarthy (2010), because of the wide acceptability for variation/synonymy among children. Furthermore, the conditioning levels provided by Anderson (2015) are exhaustive – every language has a phonology, a grammar, semantic categories, and a lexicon. Nevertheless, it will be investigated whether this parameter can also accommodate morphs that aren't allomorphs in the strict sense. Before doing this, it is useful to provide a clear definition of allomorphy, and to specify how the concept of allomorphy can be applied in this thesis; this is undertaken in the next section.

2.3 On the notion of allomorphy

Despite the appeal of the parameter 'complexity of allomorphy', it is not entirely clear what 'allomorphy' means. In the study of morphology, 'allomorphy' lacks a homogeneous definition. A canonical definition which is summarized by Faust and Lampitelli (2016) is that allomorphy represents "two or more forms in complementary distribution" conveying "the same grammatical information". Complementary distribution means that the morphs occur in mutually exclusive contexts. Nonetheless, this definition prompts the need for additional clarification regarding the precise meanings of "form" and "(grammatical) information." Other understandings of allomorphy specify the formal relationship, where allomorphs are phonological variants of one morpheme (see Booij 2005: 31; Harley 2006: 131; Lieber 2009, cited in Haspelmath 2020: 129). For example, [z] [s] [ɨz] are the phonological variants of the English plural morpheme {s}, and these variants are phonologically similar to the morpheme. This view, however, would exclude complementary morphs that are commonly regarded allomorphs but do not bear phonetic resemblance to one another, such as nominative markers *–i* and *–ka* in Korean (Haspelmath 2020: 120). A problem could arise in determining the shape of the underlying form of those allomorphs – do we assume {i} or {ka} for the underlying morpheme?

If by 'underlying form' a phonetically underspecified form is understood, such as from a generativist perspective, there is no problem to posit such a form. In this case, the feature [nominative] would stand as the abstract representation and then [i] and [ka] as surface representation. On the other hand, if we assume that morphs *and* morphemes are a pairing of form and meaning, and features like [nominative] are semantic in nature, an abstract morpheme must also have an abstract shape. Figure 2.1 shows the relation between abstract shape (morpheme), concrete shape(s) (allomorphs) and semantic features.

Abstract shape	{mente}	{s}/{z}	{?}
Concrete shape(s)	[mente]	[z]/[s]/ [ɨz]	[i]/ [ka]
Semantic Features	Adverbial	Plural	Nominative
Morpheme structure	Spanish adverbial morpheme	English plural morpheme	Korean nominative morpheme

Figure 2.1: The relation between morphemes and allomorphs as variantless morphs (Spanish), morph variants (English) and suppletive morphs (Korean).

Most linguists would agree that the best example of allomorphy in Figure 2.1 are the English plural variants, because they are phonologically similar, and an abstract *morpheme* can be stated; that is, there is a clear derivation between the morpheme and its allomorphs. Which variant is chosen to represent the abstract formal side of the morpheme depends on the analysis – for English, {s} might be chosen for being the most frequent variant; on the other hand, {z} could be chosen as formally standing between [s] and [ɨz]. In the Spanish example, there is no need to pose a general morpheme since there are no variants – the concrete shape is also the abstract shape of the morpheme. In Korean, an abstract shape cannot be stated since the shapes are not similar. Because of the prototype-like character, Hasplemath (2020) concludes that the term 'allomorph' is not very well suited for cross-linguistic comparison; instead, it could be simply replaced by the term 'morph'. All the concrete shapes in Figure 2.1 would be called morphs, no matter whether they are the only shape for a certain content, variants of one shape, or multiple shapes used for the same meaning. Haspelmath (2020) offers a distinction between *morph variants* (such as [s], [z] and [ɨz] for plural), and *supple(tive)morphs* (such as [i] and [ka] for Korean nominative). This substitution might resolve the issue of having to define the abstract shape of a morpheme. Another reason for Haspelmath's (2020: 125) concerns about employing the term 'allomorph' is that 'allomorphy' implies a process in the derivation from morpheme to allomorph; a process that is not plausible from a functionalist, usage-based perspective.

In the context of this thesis, the abstract representation of a morpheme holds no relevance. Does this imply that the term 'morph' might be a more suitable choice to encompass the complexity of lexically conditioned inflection? Not quite. The concept of 'allomorphy' is useful for two compelling reasons. First, one can argue that allomorphy specifies the complementary character of morphs: that the morphs are mutually exclusive, part of the same paradigm and have the same features. Second, this complementary distribution of allomorphy also involves a conditioning context, which is the most relevant property with regards to complexity. Treatises on allomorphs make clear that allomorphs are usually primarily classified according to their conditioning context (see Neef 2000a, b; Booij 2005; Harley 2006; Bonet and Harbour 2012), and not whether they can be derived from abstract morphemes or not. For example, Booij (2005: 31–34) says that there is a radical distinction between phonologically conditioned and non-phonologically conditioned allomorphs (which is similar to Anderson's 2015 view of allomorphy). For Bonet and Harbour (2012), a feature set "is said to exhibit allomorphy if, instead of having a unique exponent, it has two or more contextually conditioned exponents." They illustrate this relation between feature set [F], exponents[4] $\phi_{1,2}$ and the conditioning context, which can be phonological, morphological, syntactic or lexical (Figure 2.2). Figure 2.3 shows the phonologically conditioned genitive allomorphs in Kalkatungu and Figure 2.4 the plural allomorphs in English.

$$[F] \quad \Leftrightarrow \quad \begin{cases} \phi_1 & \text{Context}_1 \\ \phi_2 & \text{Context}_2 \\ \dots & \end{cases}$$

Figure 2.2: Allomorphy as a relation between semantic features, morphs and conditioning context.

[4] I will not use the term 'exponent' in this study, instead, the term 'morph' is better suited, as it is more general.

$$
\text{[GEN]} \quad \Leftrightarrow \quad \begin{cases} \text{[ku]} & \text{C\_\_} & \textit{tuat–ku} & \text{‘of the snake’} \\ \text{[ja]} & \text{V\_\_} & \textit{macumpa-ja} & \text{‘of the moon’} \end{cases}
$$

Figure 2.3: Phonologically conditioned genitive allomorphy in Kalkatungu (Blake 1969: 33, cited in Bonet and Harbour 2012: 197)

$$
\text{[PL]} \quad \Leftrightarrow \quad \begin{cases} \text{[s] [-voiced]\_} \\ \text{[z] [+voiced]\_} \\ \text{[ɪz] [+strident}^{\text{a}}\text{]\_} \end{cases}
$$

Figure 2.4: Phonologically conditioned plural allomorphy in English. [a]Includes alveolar and postalveolar fricatives.

In a way, the prefix *allo-* ‘other’ in *allomorphy* could be rather interpreted as standing for ‘other conditioning contexts’ than for ‘other forms’. Or alternatively, ‘other’ in the sense that only specific contexts allow morphs to be related to one another. Thus, [z], [s] and [ɪz] are plural allomorphs in English within a phonological context, and [s] and [en] are plural allomorphs in English within a lexical context (cf. ‘cat–s’ v. ‘ox–en’). Finally, [z], [s] and [en] can be called allomorphs that combine lexical and phonological contexts. For further purposes, the following definition of allomorphy should be provided:

> *Allomorphs are morphs that share the same semantic features and are in a complementary distribution that is conditioned by particular linguistic contexts.*

Focusing on the property of *conditioning* one can depict allomorphs as integrated within the relationship between conditioning and conditioned structure. The conditioning structure refers to the 'particular linguistic contexts' in the definition which can be phonological, morpho-syntactic or semantic; the conditioned structure of allomorphs is formal in the sense that it refers to the phonemes of a morph being changed or the entire morph ("morph variant" vs. "supplemorph" according to Haspelmath), as Figure 2.5 shows.

Conditioning structure		*Conditioned structure*
Form (phonological)	⇒	Form (phonemes of a morph)
Form-meaning (morpho-syntactic)	⇒	Form-meaning (choice of a morph)
Meaning (semantic)	⇒	Meaning (features of a morph)

Figure 2.5: Allomorphy in the broader sense. Allomorphs are form-meaning pairings whose phonological form or form-meaning pairing is conditioned (grey cells) by either semantic, morpho-syntactic or phonological structure.

Allomorphy thus excludes morphs that differ in meaning but have the same form (syncretism). In the last section, it has been argued that lexically conditioned inflection is not only complex because of allomorphy but because of syncretism. The conditioning structure – conditioned structure relationship thus shows that 'complexity of allomorphy' could be expanded to 'complexity of conditioning of morphs' while retaining the property of mutually exclusive morphs of the same paradigm.

The discussion of allomorphy shows the strengths and weaknesses to implement this concept as a variable to capture inflectional complexity. The paradigms presented at the beginning of this dissertation are certainly complex because they contain allomorphs that are conditioned, among other linguistic context, by the lexicon. The complementary distribution of these morphs and paradigms is a defining feature of lexically conditioned inflection. Nevertheless, allomorphy does not fully encompass the paradigmatic complexity of inflection; it leaves out syncretism, and – although not further discussed here – multiple exponence. Multiple exponence occurs when two morphs are used to express one category, such as the German Perfective circumfix allomorph *ge- -en* like in ***ge**-fall-**en*** 'fallen' and ***ge**-…-**t*** like in ***ge**-lieb-**t*** 'loved'. These two examples show that only one part of the circumfix is in complementary distribution, namely ***-t*** vs. ***-en***. Yet, this alternation is part of the Perfective paradigm; and the alternation between *-en*, and *-t* is as lexically conditioned as the conjugation classes in Spanish. With regards to grammatical and lexical conditioning, it would be plausible to include these other non-allomorphy types with the parameter, a parameter that would capture the complexity of *paradigmatic* conditioning in general.

Figure 2.5 shows that the conditioning contexts are classified as form (phonology), form-meaning (morpho-syntax) and meaning (semantics), which is different from the division established through Anderson's (2015) assumption: phonological (least complex), grammatical/semantic (more complex) and lexical conditioning (most complex). Because this latter division reflects complexity, these conditioning contexts will be more closely examined, and examples of conditioned allomorphs and syncretic morphs will be provided.

2.4 Relevant distinctions for conditioning of morphs

The discussion in Section 2.3 has shown that conditioning of allomorphy may approximate the complexity of conditioning of paradigms; under both views lexically conditioned inflection is considered complex. In the following, the word 'morph' will be used to encompass allomorphs, syncretic morphs and discontinuous morphs found in multiple exponence; however, the complementary and paradigmatic character of allomorphs will be implied. The first distinction that exists is between conditioned and unconditioned morphs. Unconditioned morphs are morphs that do not have a complementary distribution – "variantless morphs" (Haspelmath 2020) that are not allomorphs according to the definition provided above, or syncretic. An example is the adverbializing suffix *–mente* in Spanish, whose form and position is predictable from its own feature(s) alone ([+adverbial]). Because this morph does not have a complementary distribution that is conditioned by specific context, it does not exhibit *complexity of allomorphy*. In this regard, unconditioned morphs are less complex than morphs that are either phonologically, grammatically/semantically or lexically conditioned. Whether this general division is also reflected in speaker difficulty must be further explored.

2.4.1 Phonological conditioning

Phonologically conditioned morphs are not always considered allomorphs proper (Booij 2005; Anderson 2015). However, most of the phonological alternations to morphs bring about morphs that are in complementary distribution depending on the phonological context. This is exemplified by Ura, where the negation morph is *etw-* before vowels (1a) and *etu-* before consonants (1b):

(1) Ura (Crowley 1999: 165)

a. *Ø-etw-eni*

2sg:IMP-NEG-eat

'Don't eat it!'

b. *ir-etu-teti*

2pl-NEG-return

'Don't (you all) return!'

Etw– and *etu–* can be classified as "morph variants" according to Haspelmath (2020) because they share a string of segments, [et]. Usually, the immediate phonological environment is the reason for the formal alternation between the two variants; the monosyllabic variant *etw–* can be 'naturally' derived from the existence of an adjacent vowel, and *etu–* from the existence of an adjacent consonant. In other cases, the conditioning context must not be adjacent, as in the case of vowel harmony. In Hungarian, the allomorphs *–ok*/*–ek* alternate according to the backness of the vowel in the stem (*–ok* after back vowels, 2a; *–ek* after front vowels, 2b). Hence, the conditioning segment is not adjacent to the conditioned suffix.

(2) Hungarian

a. *olv**a**s-**o**k*

read-1sg

'I read.'

b. *lesz-**ek***

be.FUT-1sg

'I will be.'

Furthermore, the conditioning structure can be suprasegmental. Bonet & Harbour (2012) mention one example from Zapotec where the shape of the perfective suffix (*–oh*/*–ɛh*) is conditioned by the number of syllables of the stem:

(3) Tzeltal (Mayan; Walsh Dickey 1999, cited in Bonet and Harbour 2012: 222)

a. *-oh* after monosyllabic stems

*s-mah-**oh***	*s-pas-**oh***	*s-jom-**oh***
'He has hit smth.'	'He has made smth.'	'He has gathered it.'

b. *-ɛh* after polysyllabic stems

*s-majlij-**ɛh***	*s-tikun-**ɛh***	*s-maklij-**ɛh***
'He has waited for smth.'	'He has sent smth.'	'He has listened to smth.'

Phonological conditioning can not only apply to morph variants, but also to suppletive morphs[5]. This is illustrated by the choice of *–ja* vs. *–ku* in Kalkatungu, where *–ja* is used after

[5] I prefer the term 'suppletive morphs' instead of Haspelmath's (2020) neologism 'supplemorph' for the reason that 'suppletive morphs' suggests a concept that is more accessible and open to elaboration. Determining the relationship between 'supplemorphs', 'suppletive roots' and 'suppletion' is not part of this thesis.

vowels and *–ku* after consonants (Figure 2.3). Here, the two morphs do not share any segments, and as such, they can be considered suppletive morphs. Similarly, phonological features might condition an entire set of morphs, as is the case with agreement prefixes in Ingush. Here, agreement prefixes are only present when the stem is vowel-initial. In (3a–c), the verb stem {u'} agrees with the gender of the object noun phrase by prefixation of *d– b–*, *j–*. The verb stem {tieda} (3d) does not host agreement prefixes as it starts with a consonant.

(4) Agreement with gender of the object in Ingush (Nichols 2011b: 335, 398)

a.

Aaz	*dulx*	***d**-u'*
1sg.ERG	meat	**D**-eat

'I eat meat.' (Nichols 2011b: 335)

b.

Aaz	*wazhazh*	***b**-u'*
1sg.ERG	apple.pl	**B**-eat

'I eat apples.' (Nichols 2011b: 335)

c.

Aaz	*meaq*	***j**-u'*
1g.ERG	bread	**J**-eat

'I eat bread.' (Nichols 2011b: 335)

d.

Meaq	*wa-tieda/*wa-**j**ieda/*wa-**j**tieda*
bread	down-(**J**)cut.IMPF

'Cut the bread.' (Nichols 2011b: 398)

Phonological conditioning therefore affects morph variants, suppletive morphs, and entire paradigms. Conversely, the conditioning phonological structure can be segmental or suprasegmental. What unites these cases is that the conditioning structure is phonological.

2.4.2 Grammatical conditioning

Anderson (2015) characterizes the allomorphs that are conditioned by grammatical and semantic categories as more complex than phonologically conditioned allomorphs, and less complex than lexically conditioned allomorphs. For practical purposes, ‘semantic’ and ‘grammatical’ should both clustered as ‘grammatical’ since semantic features can be part of the grammar as well as morpho-syntactic features. As with phonologically conditioned morphs, grammatically conditioned morphs can be morph variants, suppletive morphs or entire paradigms. The first type represents the case where the conditioning grammatical structure is overtly expressed, morphologically or syntactically. In Supyire, Future is marked by the prefix *ŋ̀*-, a velar nasal containing a low tone (5). When a direct object is present, the segmental part of the prefix (the velar nasal) elides, however, the low tone is carried over to the object morpheme (in 5b, *kù* ‘it’). Thus, the form of the Future prefix depends on the presence of an adjacent preceding object noun phrase. This is contrasted in (5c), where there is an object noun phrase (*ù* ‘him’) which does not trigger deletion of *ŋ̀*– because it is not adjacent to *ŋ̀*–. Thus, the grammatical conditioning requires both the semantic [+object] and the formal properties [is adjacent to *ŋ̀*–] of the conditioning structure, that is, a specific form-meaning mapping unit. The Future prefix in Supyire is therefore morphologically conditioned.

(5) Supyire

a. *Wùù sí* ***ǹ****-gíí*

we FUT **FUT**-look.at

'We'll see.' (Carlson 1994: 129)

b. *Mìì sí* ***kù*** *tà*

I FUT **it.FUT** get

'I will get it.' (Carlson 1994: 129)

c. *Mìì* ***sí*** *ù lwɔ̀* ***ŋ̀****-kàn yìì á*

I **FUT** him take **FUT**-give 2pl to

'I'll take and hand (lit. give) him to you.' (Carlson 1994: 292)

In most cases, however, grammatically conditioned morphs are conditioned by a grammatical category/feature, without this conditioning grammatical structure being overt. In Lumun, a set of gender indexation morphs is prefixed to verbs only if these verbs are in their 'non-dependent' form (Smits 2017: 338) – a mood used in declarative clauses as in (6b, 6c), similar to the notion of 'finiteness'. The feature 'non-dependent' is realized together with aspect (completive, incompletive) through stem alternation. In the examples from Lumun, the dependent incompletive form of 'work' is *ɔ́rɛ́kɔ* (6a) and is used in hortative contexts; the non-dependent incompletive form of 'work' is *arɛ́kɔ* (6b) and the non-dependent form of 'work' is *ɔrɛ́kɔ̂t* (6c). Thus, there is no individual morph that marks the distinction 'dependent' vs. 'non-dependent'.

Yet, gender prefixes only occur in those forms that encode ‘non-dependent’ mood. Therefore, gender agreement morphs are not conditioned by the shape of the morph (stem) encoding ‘non-dependent’, but merely by the grammatical feature ‘non-dependent’.

(6) Lumun

a. *tʊ́rʊ́n* *ɔ́rɛ́kɔ*

1inc.HRT work.DEP.INCP

‘Let us do some work.’ (Smits 2017: 227)

b. *ɔrʊn* ***t̪**-arɛ́kɔ*

1inc **HUM.pl**-work.NDEP.INCP

‘We (INCL) will work.’ (Smits 2017: 207)

c. *ɔʊn* ***p**-ɔrɛ́kɔ̂t*

1sg **HUM.sg**-work.NDEP.CMP

‘I have worked.’ (Smits 2017: 208)

Now the reader could argue that if gender agreement morphs only occur in independent verb forms, the prefixes could also be analyzed as encoding dependency in addition to gender; i.e., *p–* would be glossed as ‘HUM.sg.NDEP’. This would mean that the conditioning category is a feature of the conditioned morph. This would represent a problem if we only limited the research to proper allomorphs – morphs that exhibit the same feature: if the morphs are interpreted being conditioned by an unrealized feature, they could be conditioned allomorphs; if the morphs are interpreted as cumulative morphs, they are no longer allomorphs since they differ in one feature. Under the broad approach pursued in this thesis, which aims to include conditioned morphs that are not allomorphs, the distinction would matter to a lesser extent. The question is whether the complexity of a morph is different between cumulative morphs and grammatically conditioned non-cumulative morphs. This question will be answered later; it could be that the complexity differs depending on the type of cumulative morphs. Provisionally, cumulative morphs that suggest the existence of two interacting grammatical categories/paradigms can be also interpreted as grammatically conditioned morphs. An interaction of paradigms with numerous features can be found for example in Kiowa, where one prefix expresses the indexation features of both subject and object referents. *gyá–* expresses 'first singular agent acting on second or third person singular patient', *gɔ́–* 'agent acts on second singular patient' and *ę̂–* 'third dual agent acts on first or third person singular patient.'

(7) Kiowa (Watkins 1984: 120)

a. *Cę̂:* ***gyá****-ɔ̨̂:*

horse **1sg.A>2/3sg.P**-give.PERF

'I gave you/him a horse.'

b. *Cę̂:* ***gɔ́****-ɔ̨̂:*

horse **A>2sg.P**-give.PERF

'We/he/they gave you a horse.'

c. *Cę̂:* ***ę̂:**-ɔ̀:*

horse **3du.A>1/3sg.P**-give.PERF

'They (du) gave me/him a horse.'

These examples show how difficult it is to distinguish between conditioned and conditioning structure with grammatically conditioned morphs. Descriptions like "the first person singular patient prefix always implies a second or third singular agent" (Watkins 1984: 119) suggest that first person singular patient morphs are conditioned by second or third person singular agent morphs. However, one could also say that *gyá–* is a second/third person singular patient morph conditioned by a first person agent. A practical solution is to compare the entire paradigm and see which formal part of the morph is more invariant depending on several combinations; this invariant shape could be determined as the conditioned structure. By doing so, one can include cumulative morphs into the tripartite distinction of complexity based on conditioning which was established on the base of allomorphy by Anderson (2015).

2.4.3 Lexical conditioning

Lexically conditioned morphs comprise a context that is very narrowly specified, namely lexemes. Lexical conditioning is distinguished from phonological and grammatical conditioning in that the latter two types comprise features/morphs of a closed class. The conditioning structure in lexically conditioned allomorphs always refers to morphs that belong to an open class. While the distinction between 'open' and 'closed' might be artificial, this distinction relates to the number of elements falling into a specific class, and this number, by being small or large, impacts the degree of complexity. The conditioning structure of lexically conditioned allomorphs is part of a paradigm with countless members, namely the lexicon. However, not every lexeme directly conditions allomorphy. Lexically conditioned allomorphs appear in word forms that can be called ‘irregular’. Irregular allomorphs can be conditioned by either a single lexeme or a group of lexemes. For example, the English plural allomorph *–en* is conditioned by the lexical morph *ox*; that is, *–en* ‘plural’ appears only in combination with *ox*:

[PL] ⇔ [en] ox *ox-en*

The conditioning structure of lexical conditioning is always *morphological* and cannot be merely semantic or phonological. That is, the occurrence of the suffix *-en* is defined as pertaining to the specific semantics ‘castrated adult male cattle’ with the specific form [ɒks]. The semantics alone do not predict allomorphy, as exemplified by the synonym *bullock* which is *bullock–s* in the plural; neither does merely the phonology, as exemplified by other nouns ending in [ɒks] like *fox*, which is *fox–es* in the plural. Lexically conditioned allomorphs might apply to several lexemes, as is the case for irregular past forms in English. According to the *Grammar for teachers* (DeCapua 2017: 401), there are nine verbs that form their past by stem vowel alternation from [ɪ] (present) to [æ] (simple past):[6]

(8) Verbs in English with present [ɪ] to simple past [æ] stem alternation ‘drink–drank’

‘ring–rang’

[6] See Bybee & Slobin (1982) for a more refined classification based on phonological schemas across irregular verbs.

'shrink–shrank'

'sing–sang'

'sink–sank'

'spring–sprang'

'stink–stank'

'swim–swam' (DeCapua 2017: 401)

This group of verbs is however neither generalizable on the semantic level (there are both transitive and intransitive verbs inside and outside this group) nor fully on the formal level (having an [ɪ] in the nucleus of the present from and some nasal element on the coda does not condition [æ] in the simple past; see *wink – winked*). As such, the allomorph C[æ]C is conditioned by each lexeme pertaining to this group, not by the whole group as an abstraction from its members. If these groups have a multiplicity of lexemes triggering specific allomorphs the term 'lexical class' is preferred. These classes are distinguished by recurring allomorph sets. Compare the different Spanish allomorph sets for person/number/mood inflection (Table 2.1).

	a–class {am} 'love'		**e-class {com} 'eat'**		**i-class {viv} 'live'**	
	indicative	subjunctive	indicative	subjunctive	indicative	subjunctive
1sg	am-**o**	am-**e**	com-**o**	com-**a**	viv-**o**	viv-**a**
2sg	am-**as**	am-**es**	com-**es**	com-**as**	viv-**es**	viv-**as**
3sg	am-**an**	am-**e**	com-**e**	com-**a**	viv-**e**	viv-**a**
1pl	am-**amos**	am-**emos**	com-**emos**	com-**amos**	viv-**imos**	viv-**amos**

2pl	am-**áis**	am-**éis**	com-**éis**	com-**áis**	viv-**ís**	viv-**áis**
3pl	am-**an**	am-**en**	com-**en**	com-**a**	viv-**en**	viv-**an**

Table 2.1: Indicative and subjunctive present forms of *amar* 'love' *comer* 'eat' *vivir* 'live' in Spanish.

Verbs of the a-class encode person/number with morphs beginning with [a] in indicative (except 1sg) and morphs beginning with [e] in subjunctive. Verbs of the e-class encode these features with morphs beginning with [e] in indicative (except 1sg) and morphs beginning with [a] in subjunctive. Lastly, verbs of the i-class encode these features with morphs beginning with [e] for 2sg, 3sg, 3pl, with [i] in 1pl and 2pl and [o] for 1sg in indicative and morphs beginning with [a] in subjunctive. These patterns are predictable once the class is known; one form of the paradigm can predict another form. However, the *choice* of a certain set is not predictable in the sense that they are conditioned by lexical morphs. A paradigm is associated with a group of lexical morphs, but this group, again, is neither predictable from the phonological nor semantic features alone exhibited by its members (see example 9).

(9) Excerpt of verbal stems pertaining to different conjugation classes in Spanish

a-class: /am/ 'love' /jam/ 'call' /miɾ/ 'look' /estudi/ 'study' /dese/ 'desire' (...)

e-class: /lam/ 'lick' /le/ 'read' /ka/ 'fall' /respond/ 'respond' /bar/ 'sweep' (...)

i-class: /bib/ 'live' /kompaɾt/ 'share' /un/ 'join' /abɾ/ 'open' /o/ 'hear' (…)

The triple period means that the list is not exhaustive. All classes contain various semantically (e.g., transitive/intransitive) and phonologically (mono- and polysyllabic; closed and open final syllables) diverse members, which makes it impossible to generalize a pattern. This means that the allomorphs are not conditioned by their lexical class, but indirectly by the lexical morphs that occur in these classes. The terms a-/e- or i-class are generalizations of the *conditioned* structure, not of the *conditioning* structure. The information relating classes to lexemes or vice versa must be mentioned in the grammar or the lexicon (usually in the latter). Therefore, the list of lexemes is always implied when capturing the conditioning structure, and can be illustrated as follows ("...}" indicates that the list is not exhaustive):

[INDICATIVE.NON1sg] ⇔ [a] a–conjugation: {am; tom; llam; mir; dibuj; estudi; ...}

Parallel to grammatical conditioning, it is at times not easy to keep conditioned and conditioning structure apart. That is – to what extent is a morph *lexically conditioned* or part of the lexical morph? In the Spanish examples, segments of the morphs like [a] and [e] might be also interpreted as part of the stem, as stem vowels alternating for mood/person and not as part of mood/person suffixes that are conditioned by the lexical morph. How does it apply to more prototypical cases of stem alternation, as also exemplified by English verbs? In those cases, the segments that alternate across different word forms are part of the stem and express a set of features. In Dumi (Sino-Tibetan, Nepal), stems have one to four forms depending on their lexical class and inflectional features expressed. The non-negated simplex conjugation of the stem morphs *buts/bus-bo?* 'shout, cry; crow (of a cock)' is presented in Table 2.2.

	Non-past	Past	Alternation
1sg	bus-tə	buts-ə	*bus/buts*
2sg	bus-ti	buts-i	
1du.exc	bus-tɨ	buts-ɨ	
1pl.inc	boʔ-kti	boʔ-ki	*boʔ*
1pl.exc	boʔ-kta	boʔ-ka	
2sg	a-bus-ta	a-buts-a	*bus/buts*
2du	a-bus-ti	a-buts-i	
2pl	a-bus-tini	a-buts-ini	
3sg	bus-ta	buts-a	
3du	bus-ti	buts-i	
3pl	ham-bus-ta	ham-buts-a	

Table 2.2: Non-negated simplex conjugation of the stem buts/bus-boʔ ‘shout, cry; crow (of a cock)’ Van Driem (1993: 98)

	Non-past	Past	Alternation
1sg	ləs-tə	lənts-ə	
2sg	ləs-ti	lənts-i	*ləs/lənts*
1du.exc	ləs-tɨ	lənts-ɨ	
1pl.inc	lo:-kti	ləʔ-ki	*lo:/ləʔ*
1pl.exc	lo:-kta	ləʔ-ka	
2sg	a-ləs-ta	a-lənts-a	*ləs/lənts*
2du	a-ləs-ti	a-lənts-i	
2pl	a-ləs-tini	a-lənts-ini	
3sg	ləs-ta	lənts-a	
3du	ləs-ti	lənts-i	
3pl	ham-ləs-ta	ham-lənts-a	

Table 2.3: Non-negated simplex conjugation of the stem ləs/lənts- lo:/ləʔ 'come out, emerge' (Van Driem 1993: 99)

Compare Table 2.2 and 2.3: For the verb 'shout, cry', there are three stem allomorphs: In non-past, *bus* is used for all person features except for first plural which is expressed by *boʔ*; in past, *buts* is used for all person features except for first plural which is expressed by *boʔ*. For 'come out, emerge', there are four stem allomorphs: *ləs* for all persons in non-past, except *lo* for first plural; *lənts* for all persons in past except *ləʔ* for first plural. These alternations are not predictable from the phonology, since /bus/ and /ləs/ have the same coda; nothing about /bus/ and /ləs/ suggests that the past tense stem allomorphs will be /buts/ and /lənts/. Instead, the shape as well as the number of stem allomorphs depends on the conjugation class of the stem. Van Driem (1993: 98) subsumes the pattern of 'shout, cry' to 'intransitive conjugation 3' and the pattern of 'come out, emerge' to 'intransitive conjugation 5.'

What is however different from the Spanish examples is that the alternating segment is clearly part of the stem – these segments (*us* and *oʔ* for 'shout, cry' and *s* and *ʔ* 'come out, emerge') are both grammatical and lexical morphs – grammatical because they alternate for grammatical features, and lexical because their shape is idiosyncratic and is not shared by other lexemes. The lexical information in 'come out, emerge' is expressed by onset /l/, the nucleus /ə/ and the coda /s/. /ləs/ alternates to /lo:/ in first plural; here, it would be non-intuitive to say that /o:/ is an allomorph for first person plural, whereas /əs/ for non-first-person-plural. Rather, parts of the stem morph are conditioned by the grammatical features, not the other way around.

Moreover, there is the case where lexically conditioned stem alternation is pervasive across all word forms but does not fall into specific classes. In Navajo, every stem alternates for certain inflectional categories. The choice of these categories as well as the shape of the alternation is conditioned by the stem itself. This is exemplified by two stem sets in Tables 2.4 and 2.5.

<table>
<tr><th></th><th>Imperfective</th><th>Optative</th><th>Perfective</th><th>Iterative</th><th>Future</th></tr>
<tr><td>Momentaneous</td><td colspan="2">lííd</td><td rowspan="2">lid</td><td rowspan="2">li'</td><td rowspan="2">lił</td></tr>
<tr><td>Durative</td><td colspan="2">lid</td></tr>
<tr><td>Neuter (stative)</td><td>lid</td><td></td><td>lííd</td><td></td><td></td></tr>
<tr><td>Repetitive</td><td>li'</td><td></td><td></td><td></td><td></td></tr>
</table>

Table 2.4: Stem paradigm of *lid* 'smoke, burn' (summarized and rearranged from Young, Morgan and Midgette 1992: 370-71)

<table>
<tr><th></th><th>Imperfective</th><th>Optative</th><th>Perfective</th><th>Iterative</th><th>Future</th></tr>
<tr><td>Momentaneous</td><td colspan="2">chíid</td><td>chid</td><td rowspan="2">chi’</td><td rowspan="2">chił</td></tr>
<tr><td>Continuative/Reversative</td><td colspan="3">chid</td></tr>
<tr><td>Repetitive</td><td>chid</td><td colspan="4"></td></tr>
</table>

Table 2.5: Stem paradigm of *ch’id* ‘act abruptly with hands’ (Young, Morgan and Midgette 1992: 83)

Stems in Navajo are always monosyllabic, and their onset expresses only lexical information, whereas the nucleus and coda express lexical and grammatical information. The difference between the stem paradigm ‘smoke/burn’ and ‘act abruptly with hands’ is that the former allows ‘neuter’ and ‘durative’ forms, whereas the latter does not have these forms but instead allows ‘continuative/reversative’ forms. The set choices are due to the specific semantic differences between ‘smoke/burn’ and ‘act abruptly with hands’, but the semantic differences cannot be generalized for all verbs. That is, it is not predictable from the semantics of a stem morph whether it allows ‘neuter’ ‘durative’ or ‘continuative/reversative’ forms. In addition, the shape of these morphs is not predictable – Repetitive Imperfective in ‘smoke/burn’ is expressed by the rime /i’/ whereas in ‘act abruptly with hands’ it is expressed by /id/. These peculiarities are pervasive in Navajo stem alternation; and the number of categories as well as their formation makes it hard to posit clear-cut lexical classes. They are like stem alternations in English, with the difference that every verb in Navajo exhibits stem alternation for some grammatical categories.

That this type of stem alternation is not merely the expression of lexical plus grammatical material, but is, in addition, lexically conditioned, is understandable when comparing other types of stem alternation. In Jingulu (Mirdni, Northern Territory), there is regressive vowel harmony. “Certain affixes containing high vowels (/i/ or /u/) trigger a raising of the final vowel of the stem, if it is low (/a/), to /i/” (Pensalfini 2002: 561–562). For example, the morph *ngamurla* (big) becomes *ngamurli* when the suffix /-rni/ (feminine) is added: *ngamurl**i**rni* ‘big’ (feminine) (Pensalfini 2002: 562). Here, the choice of two lexical allomorphs *ngamurla/ngamurli* is phonologically conditioned by the vowel of the following morpheme. Like grammatical morphs, lexical morphs can be phonologically, grammatically or lexically conditioned. An example of a lexical morph being conditioned by another lexical morph is the English compound *shepherd*, [ʃɛpəɹd] where [ʃɛp] is a morph variant of [ʃip] ‘sheep’ occurring only in combination with the lexical morph *herd* [əɹd]. In other analyses, the word *shepherd* is monomorphemic. In sum, lexical morphs can be both conditioning and conditioned structure.

Finally, lexically conditioned allomorphs can manifest themselves in stem suppletion. Consider the forms of ‘be’ in English: I *am*, you *are*, s/he *is*; here, there is no segment that remains constant, and as such, there are no alternating or invariant segments. With suppletion, it is practically impossible to say whether the lexical information of the stem or the grammatical information of the stem are conditioning categories since there isn’t a contrast between alternating and non-alternating segments that would indicate a conditioning-conditioned (as is the case for stem alternation). This is even more evident when the grammatical categories expressed are more derivational, like singular/plural stem allomorphy in Northern American languages (Mithun 1988), such as in Kiowa, where *ét* means ‘big (singular)’ and *bîn* means ‘big (dual/plural)’. Bonet & Harbour (2012: 215-2016) conclude that “the feature [± singular] does not condition allomorphy of *ét* versus *bîn* ‘big’, but *ét* (or *bîn*) is the realization of a version of the stem of [‘]big[’] that selects for a [+singular] or [-singular] complement” (cf. Watkins 1984: 154). Another interpretation is that stem suppletion for grammatical features is the opposite of lexically conditioned grammatical affixes: here, the choice of a stem allomorph is conditioned by grammatical features. Despite this ambiguous status of stem suppletion (see Table 2.6) – parallel to cumulative morphs (where one morph expresses features of two grammatical categories, as shown in 2.4.2) – stem suppletion and lexically conditioned grammatical morphemes have in common that they both entail additional features. Table 2.6 summarizes the relation between lexical morphs and lexical conditioning.

	Conditioning context	**Conditioned context**
Lexically conditioned affixes	stem	affix
Lexically conditioned stem alternation	feature + stem	stem segment(s)
Root suppletion	feature	stem

Table 2.6: Types of lexically conditioning allomorphs (grey) and stem suppletion (white)

In all instances of lexical conditioning, the conditioning structure is a lexical morph (i.e., stem). The clearest example of lexical conditioning is when the conditioning and conditioned structure are separate morphs. With stem alternation, a part of the lexical morph is conditioned by grammatical features; but, as Dumi and Navajo show, these features are themselves lexically conditioned in how they impact the stem. As such, lexically conditioned stem alternation represents a case of bi-conditioning where a lexical morph is conditioned by non-lexical features; but the instantiation of these features depends on the same lexical morph. Stem suppletion can be analyzed as being conditioned by non-lexical features or as expressing these non-lexical features cumulatively.

The abundance of cumulative morphs, stem alternation, suppletion and syncretic morphs in lexically conditioned inflection shows the need for these morphs and their paradigms to be integrated into the tripartite complexity distinction of Anderson (2015), even if conditioning or the directionality of conditioning cannot be determined easily.

2.4.4 Multiple conditioning

The examples provided in this section illustrate distinct types of conditioning, but the reader might have noticed that morphs do not have to be solely phonologically, grammatically or lexically conditioned. Indeed, there can be a multiplicity of conditioning contexts that need to be listed in the description of allomorphy, syncretism and cumulation. One subtype of lexical conditioning is inherently bi-conditioned, namely stem-alternation (see Table 2.6). Furthermore, an allomorph can be a morph variant or a suppletive morph, depending on which allomorph it is associated with. One example from English is allomorphy in nominal plurality, which is lexically conditioned regarding the choice of allomorph but phonologically conditioned for morph variants of *-s*.

[pl] ⇔
[ən] ox
_[ɪ]_[rən] child
Largest group of nouns ⇔
[s] [–voiced]_
[z] [+voiced]_
[ɪz] [+palatal/alveolar] [+fricative]

Figure 2.6: Multiple conditioning of plural allomorphy in English

Another example of multiple conditioning can be found in Oneida. Verbal indexation prefixes are conditioned by multiple grammatical categories such as subject and object features (grammatically conditioned), the shape of the following morph (phonologically conditioned) and the stem morph (lexically conditioning).

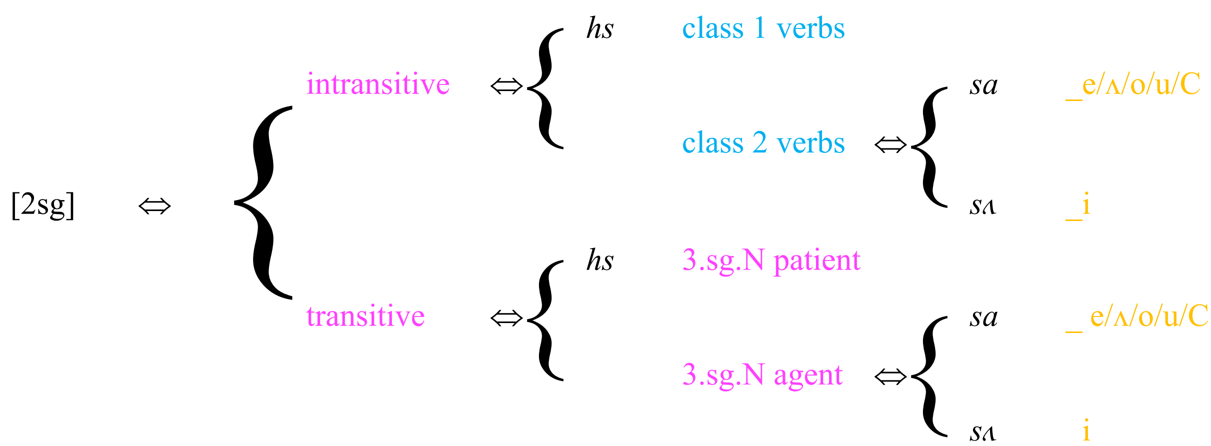


Figure 2.7: Multiple conditioning of second person singular allomorphs in Oneida (Cf. Abbott 2000: 22–31). Conditioning types are illustrated in colors: (phonological, grammatical, lexical conditioning).

Whereas transitive allomorphs are grammatically and phonologically conditioned, intransitive allomorphs are also lexically conditioned. The choice of *hs*– vs. *sa*–/*sʌ*– depends on the lexical morph they are attached to. Abbot (2000) finds some generalizations, but they are not exhaustive: “Most dynamic verb stems require subjective prefixes [here: *hs*-], except when the stative suffix is added, while others require objective prefixes [here: *sa*–/*sʌ*–]” (Abbot 2000: 38). This means that despite the tendency to predict the choice of prefixes in intransitives, there are exceptions which can be only accounted by lexical idiosyncrasies. On a higher level, allomorph choice is primarily conditioned by transitivity, i.e., grammatically conditioned. This interplay suggests that morphology in Oneida is extremely complex, as Koenig and Michelson (2015) conclude.

2.4.5 Co-occurrence in constructions

What has not been discussed is whether morphs conditioned by distinct structures can appear together in one construction. In principle, there is no reason to assume otherwise. Conditioning levels are not mutually exclusive or need to co-occur. A language with constructions exhibiting morphs each conditioned by a different linguistic level is Ura (Austronesian; Crowley 1999). Consider the following examples (10 a–d):

(10) Ura (Crowley 1999: 165)

a. *y*-*etw*-*arufa*

3sg.DPT-NEG-sing(DPT)

'S/he did not sing.'

b. *c*-*etwi*-*narufa*

3sg.FUT-NEG-sing(FUT)

'S/he will not sing.'

c. *y*-*etw*-*arap*

3sg.DPT-NEG-sit(DPT)

'S/he did not sit.'

d. *c*-*etw*-*adap*

3sg.FUT-NEG-sit(FUT)

'S/he will not sit.'

Ura shows different types of conditioning for different morphs:

i. Phonological conditioning (orange): *etw-* (negative) is used before vowels (10a, c, d); *etwi-* (negative) before consonants (10b) (Crowley 199: 165).

ii. Grammatical conditioning (magenta): *y-* (3rd person) is used in Future (10a, c); *c-* (3rd person) in Distant past (10b, d). (Crowley 1999: 165).

iii. Lexical conditioning (blue): Different stem segments are conditioned by the features Distant past and Future; the alternation for 'sing' applies to the initial syllable [na]/[a], whereas for 'sit', it applies to the middle consonant [r]/[d] (Crowley 1999: 165; 151–154).

After describing the different (sub)types of conditioned morphs, the question of their correlation with degrees of morphological complexity should be further elaborated. Section 2.5 will discuss this question in regard to the different measures of complexity presented in Section 2.2.

2.5 Complexity correlates across conditioning types

Throughout Section 2.4, it has been stated that allomorphy adds, in principle, to morphological complexity. From a description-length perspective, allomorphy implies the specification of information on both the conditioned and a conditioning structure. The definition of allomorphy as primarily conditioned morphs in complementary distribution with the same features reveals another problem, namely that allomorphy is most clearly defined when conditioning and conditioned structure are segmentally separate from one another. The fact that this distinction is not always straightforward shows that there is a continuum between allomorphy and cumulative morphs. On the one end of the continuum, an invariant set of morphs is used in complementary contexts, such as gender agreement for non-dependent verbs, as shown in Lumun (6a–c). On the other end of the continuum, morphs do not share any segment that would suggest an association with a specific feature, such as between *gyá* (1sg.A>2/3sg.P) and *ę̂:* (3du.A>3sg.P) in Kiowa. These two morphs share the information '3rd person patient', but no segment can be associated with it. Intermediate cases are the ones where only some segments in one morph can be associated with certain features. For example, the contrast between *gyá* (1sg.A>2/3.P) and *gɔ́* (A>2sg.P) suggests that /g/ expresses second person patient (7a–b). Nevertheless, this should not speak against integrating *complexity of allomorphy* as a parameter of cross-linguistic analyses on complexity, quite the opposite: by integrating both clear cases of allomorphs and cumulative morphs it is possible to incorporate into the investigation another parameter from Anderson (2015), namely *complexity in the realization of individual elements* – in this case, allomorphs and cumulative morphs are related to one another, and could be grouped as morphs that are more complex than variantless morphs. Furthermore, syncretism, which Anderson (2014: 22) classified under the parameter *complexity of inter-word relations*, may also be more complex than variantless morphs. The question is whether cumulative morphs and syncretism can be also more complex when integrated in a parameter that establishes complexity levels based on conditioning, such as *complexity of allomorphy.* Therefore, this section needs to examine not only whether Anderson's (2015) intuition is correct about the correlation of these conditioning types to complexity, but also if his idea holds true once morph types that are complex according to the other two parameters (*complexity in the realization of individual* *elements*, *complexity of inter-word relations*) have been subsumed under *complexity of allomorphy*, expanding the parameter to *complexity of conditioning of morphs.* After all, the aim is to find viable variables for evaluating all cases of inflectional complexity against lexically conditioned inflection. In the following section, I will investigate whether a tripartite distinction on phonological, grammatical and lexical conditioning reflects linguistic complexity; this will be done with regards to description length (4.1), entropy (4.2), and finally – the most relevant perspective for this dissertation – with regards to user difficulty (4.3).

2.5.1 Complexity based on description length

The reason to start with description length is that it can quantify complexity; the question is whether there is some quantitative dimension behind the tripartite distinction between phonological, grammatical, and lexical conditioning that may affect speaker difficulty. As said in the introduction, description-based assessments of complexity might not be unrelated to speaker-based assessments, especially when the language is learned by speakers utilizing grammatical descriptions.

Allomorphy itself can increase the description length of a paradigm in the case that it multiplies the number of morphs available to express specific features. For example, the existence of three conjugation classes requires a longer description than one conjugation class for a given number of features. This also happens if phonological rules are involved in a paradigm: the more phonological rules, the more allomorphs a paradigm could have. However, it is not fully clear whether description length increases based on the level of conditioning (phonology; grammar, lexicon). In order to know if complexity of allomorphy correlates with description length, one would have to demonstrate that phonological conditioning requires a shorter description than grammatical conditioning, and lexical conditioning more than phonological and grammatical conditioning. Given that languages differ widely in the inventory size of paradigms, number of phonological rules, grammatical categories and number of lexemes, it is very hard to make a direct connection between level of conditioning and description length. Nevertheless, generalizations based on the *limit* of the number of conditioning elements within one level *in relation* to the limit of the number of conditioning elements in another level may be given.

If we focus on phonologically conditioned morphs, it is usually the case that the different formal variants are conditioned by abstract *features* of the phonological environment. For example, the English plural allomorphs [s] [z] [ɪz] are conditioned by the following features: [±voiced] [±strident]. The description of these morph variants entails listing three formal realizations, two conditioning features as well as mentioning the relation between the conditioning contexts and the realizations. Not all combinations are possible – instead, we have three rules:

1. [s] is used after voiceless segments that are not strident fricatives.
2. [z] is used after voiced segments that are not strident fricatives.
3. [ɪz] is used after strident fricatives.

This represents a manageable number of rules associated with the conditioning context. Morphs that are merely phonologically conditioned require a short mention of when to choose which variant. In principle, the description length of phonologically conditioned morphs is always shorter than the number of features of the phonological and prosodic inventory of a language. In fact, the number of rules associated with the conditioning context (phonemes, features) should be always equal[7] to the number of allomorphs, since every phonological conditioning implies that there is always a change in the shape of a morph.

With grammatically conditioned morphs, the relation is very similar: conditioning contexts are usually grammatical categories/features or single morphs, and the number of categories conditioning a set of morphs is, in principle, never larger than the sum of all grammatical categories in one language, which is a limited set. Compare the set of articles in German (Table 2.7):

<table>
<tr><th></th><th>Masculine (sg.)</th><th>Neuter (sg.)</th><th>Feminine (sg.)</th><th>Plural</th></tr>
<tr><td>Nominative</td><td>der</td><td rowspan="2">das</td><td colspan="2" rowspan="2">die</td></tr>
<tr><td>Accusative</td><td>den</td></tr>
<tr><td>Genitive</td><td colspan="2">des</td><td colspan="2">der</td></tr>
<tr><td>Dative</td><td colspan="2">dem</td><td>der</td><td>den</td></tr>
</table>

Table 2.7: Article morphs in German

Cumulative morphs like German articles, which are monomorphemic and manifest two grammatical dimensions (person/number and case) can be also read as morphs that are conditioned by either dimension: *den* can be interpreted as one accusative morph that is conditioned by the feature [masculine] or a masculine morph conditioned by the feature [accusative] (or alternatively, a plural morph conditioned by a dative feature or vice-versa). This alternative

[7] Equal in the number of rules, not in the number of features that specify a rule. Of course, many features in a rule increase the description length, but at a certain degree of specificity, it would be more convenient (and shorter) to list the single phonemes. For example, the minimal description length for [ɪz] is "[+strident]_"; yet mentioning the single phon(emes) instead of abstract features does dramatically increase the description length: "[ɪz] is used after [s, z, (t)ʃ, (d)ʒ]"

interpretation matters for the description since it is sufficient to describe either the rows or the columns. In both cases, the number of morphs (*der*, *den*, *des*, *dem*) is never larger than the number of conditioning features (nominative, genitive, dative, accusative). However, where grammatically conditioned paradigms differ from phonologically conditioned paradigms is in syncretism. There are not only several form variants for one single meaning, but also several meaning variants for one single form. The same form can be used in the same paradigm (*die* can be feminine singular nominative or accusative) or across paradigms (*die* can be feminine singular nominative or plural accusative). Syncretism reduces the number of morphs but does not reduce the number of features. Grammatically conditioned paradigms are therefore potentially more complex than phonologically conditioned paradigms (given the same number of cells in the paradigm) in terms of description length since the same morph might be conditioned by more than one feature/category; thus – the number of conditioning elements is potentially larger than the number of morphs. But how much larger? In principle, it depends on the number of syncretic forms. If all the forms are syncretic (one syncretism), there is no paradigm and therefore there would not be a need to describe a grammatical category since it is not distinguished by different forms in different features. For example, one could say that the Masculine, Feminine and Neuter forms of the German plural article are all syncretic – *die* in Nominative/Accusative. But if they are all syncretic, one can simply omit the gender distinctions from the description, as done in Table 2.7. This reduces the number of syncretic forms.

This means that the number of the conditioning features for one grammatically conditioned morph (allomorph or cumulative morph) is always *smaller or equal* than the number of morphs of the largest morph set conditioned by the same category (in Table 2.7: case/gender/number). In the case of German, the largest morph set for articles is for the feature [masculine] with 4 different forms; this morph set represents the leftmost column in Table 2.7. The feminine morph set has 2 morphs (die, der) that are conditioned by 4 features [nominative, genitive, dative, accusative], thus conditioned by an equal number of features as the masculine set. The number limitation of conditioning features for grammatically conditioned/cumulative morph paradigms derives from the limitation of the number of the largest morph set of the paradigm. This means that the description length of a grammatically conditioned paradigm is potentially larger than the description length of a phonologically conditioned paradigm but not exceedingly larger since it depends on the number of conditioning features of the conditioning category. And the number of possible features is limited because the number of grammatical features is 'finite' – i.e., easy to enumerate.

With lexically conditioned paradigms, the relation between number of morphs and number of conditioning features is radically different to the other two types. The number of morphs is almost always smaller than the number of lexemes conditioning them, especially when there are lexical classes, in which case groups of numerous lexemes condition a specific morph. Since the classes/groups of lexemes cannot be generalized under a semantic or phonological feature, the number of conditioning elements in a group is equal to the number of lexical morphs that are classified into this group. Unlike in phonological and grammatical conditioning, there is, in principle, no inherent limit to the number of conditioning elements in lexical conditioning, since lexical morphs are part of an 'open' paradigm.

With a lexically conditioned inflectional paradigm, such as the past tense allomorphs of English, the shortest description depends on the most productive pattern. For example, the shortest description of the verbs that form their simple past by C[æ]C is to enumerate the lexemes that carry this allomorph in the simple past, as done in (8) (Section 2.4.3). With a (more) productive allomorph, the shortest description would be to list the verbs that do *not* have this allomorph. The number of English verb forms that have *–ed* for simple past is (in principle) infinite; and the number of the ones that do not have an *–ed* suffix is the sum of all irregular verbs in English, about 60 (DeCapua 2017: 401–403).

The most ‘complex’ allomorphs in terms of description length are therefore the ones that are associated with lexical classes containing a great number of lexemes, but which are not clearly productive, such as in Spanish (see Table 2.1); or the non-classifiable, but pervasive stem alternation patterns as shown in Navajo (see Tables 2.4 and 2.5). If there is a productive class, then the allomorph of this productive class entails a shorter description than the allomorphs of non-productive classes, since the productive allomorph is simply defined by exclusion of the other classes that were already defined by listing their lexemes. In other words, the shortest description for the application of the English past tense allomorph *–ed* would be to write ‘all verbs except the ones of the groups listed above’. Similarly, less complex are groups of verbs/classes that contain only few members. This can be exemplified in Figure 2.8.

$$\text{[Simple Past]} \quad \Leftrightarrow \quad \begin{cases} \text{[t] bend, build, lend} \\ \text{[æ] begin, drink, ring, shrink, sing, sink, spring, stink, swim} \\ \text{(...) (other irregular groups)} \\ \text{[(ɪ)d] all verbs except the ones of the groups listed above.} \end{cases}$$

Figure 2.8: Description of allomorphy groups for English simple past.

Potentially, the number of conditioning elements is smaller or equal than half the number of lexemes hosting these allomorphs – in the case that there are only two classes and they both exhibit an equal number of lexemes. In any case, lexically conditioned allomorphs require a larger description than phonologically or grammatically conditioned allomorphs because for every allomorph there are several contexts where they can appear (like syncretisms discussed above). This is attributed to the 'open' character of lexeme classes like verbs, nouns and adjectives. An exception to lexically conditioned allomorphs being always more complex than phonologically or grammatically conditioned allomorphs is when there are allomorphs conditioned by only one lexeme. In English, there is no first person singular agreement except with the verb 'be'. Thus, 'Ø' and 'am' are the only allomorphs (if we count zero-realization as a morph) for [1sg], conditioned by the lexeme 'be'. Therefore, one must be cautious to not generalize the complexity properties associated with lexically conditioned allomorphy for these kinds of 'single' irregulars. Fortunately, grammars usually have a list of irregulars, and single irregulars could be excluded from the investigation. Table 2.8 summarizes the different generalized degrees of description-length-based complexity across the three types of conditioning contexts, in contrast to no allomorphy.

Allomorphs	**Description**	**Add. morph. complexity**
No conditioning	No additional length	None
Phonological conditioning	$N_A + N_{CE}$ ($N_{CE} = N_A$); $N_{CE} < \sum_{\text{Phonological Elements}}$	Low to middle
Grammatical conditioning	$N_A + N_{CE}$ ($N_{CE} \geq N_A$); $N_{CE} < \sum_{\text{Grammatical Elements}}$	Middle
Lexical conditioning (except single irregulars)	$N_A + N_{CE}$ ($N_{CE} \geq N_A$); $N_{CE} \leq 1/2\sum_{\text{Lexical Elements}}$ $1/2\sum_{\text{Lexical Elements}} < \sum_{\text{Grammatical Elements}}; \sum_{\text{Phonological Elements}}$	Middle to high

Table 2.8: Correlates of allomorphy with description-based complexity. N_A = Number of Allomorphs; N_{CE} = Number of conditioning elements; $\sum$ = sum.

In conclusion, description-length based measures can provide an explanation on why lexically conditioned morphs are more complex than grammatically conditioned morphs, and grammatically conditioned morphs more complex than phonologically conditioned morphs. However, the cline represented in Table 2.8 should be read with caution since description length primarily is based on the number of rules and inventory size of a paradigm. Conditioning itself does not determine complexity, but paradigms conditioned by different levels of language involve different rules and sets that can vary in size and thus in the description. Moreover, what has not been addressed in this section is how different types of morphs (derivational vs. inflectional) relate to description-based measures. Since the question raised at the beginning of this dissertation specifically targeted lexically conditioned *inflectional* paradigms, one would have to determine whether the number of lexemes conditioning derivational morphs is similar to the number of lexemes conditioning inflectional morphs.

2.5.2 Complexity based on entropy

Entropy-based measures of complexity can be applied to systems that are not fully predictable. A system has zero entropy if an item can be fully predicted based on a given context. This system predictability might indeed be reflected in users, as predictability in guessing or memorizing forms depending on other forms. Thus, entropy-based measures might be related to user-based measures of complexity. Predictability is dependent on the context. The narrower the context, and the fewer items there are to choose, the more predictable a specific item is. Thus, entropy measures, like description length measures, are always dependent on context and its size, as well as the number of the items. With regards to paradigmatic complexity, the context is the paradigm and the lexemes pertaining to the paradigm, and the items are the cells of the paradigm that can be filled with morphs. Thus, one can calculate the entropy (and thus complexity) of a specific morph in a specific paradigm for a specific lexeme given a specific set of neighboring cells of the same paradigm or given a specific lexeme for different lexemes.

For example, the cell for the form 'simple past' in English can be calculated based on the neighboring cells of the tense paradigm (present, simple past, perfective), in relation to different allomorphs/strategies to mark past tense (*-ed*, *C[æ]C*, *C[ɪ]C*) and also in relation to different lexemes (wink, ring, hit). One could then determine that the predictability for the strategy *–ed* to mark past tense for any given lexeme is higher than the strategy *C[æ]C*, such as in *ring*. Including frequency to the system would lower the predictability for choosing *-ed*, since verbs that form the past tense with another strategy tend to be quite frequent, such as *be*, *do* or *go*.

The entropy of a specific cell in the paradigm, namely the simple past form of verbs, increases when the number of allomorphs is higher, and these allomorphs can be chosen by a larger number of lexemes. According to Ackerman & Malouf (2013: 439), the highest entropy of a paradigm occurs when there are evenly-distributed conjugation classes with equal numbers of lexemes in them, such as in Spanish or Italian. The English past tense system has a lower entropy since the past tense suffix *–ed* is highly productive and only about 60 lexemes do not choose this suffix.

How do entropy measures then relate to the distinction between phonologically, grammatically and lexically conditioned allomorphs? As noted above, the predictability for choosing a specific allomorph is dependent on the fact that the variable 'lexeme' is undefined. We can see an increase in entropy for lexically conditioned allomorphs only by investigating the predictability of a specific allomorph given *any* lexeme, while fixing the category 'simple past'. System of morphs that are merely grammatically conditioned would always be predictable in a context that fixes the knowledge of all the other contexts (phonological, syntactic, semantic, grammatical) but leaves only the lexeme undefined. Vice versa, a system of morphs that are merely phonologically conditioned would also always be predictable in this context. Thus, the complexity that lexically conditioned morphs add to a paradigm depends on a specific context in which the lexeme context is unknown, but the grammatical, phonological context is known, as well as the set and number of allomorphs. Table 2.9 summarizes the potential entropy of a given specific set of allomorphs in a specific grammatical and phonological context but given an unspecified lexical context.

Allomorph system	**Entropy**	**Add. morph. complexity**
No conditioning	Zero (allomorphs are predictable from the phonological and grammatical context alone)	No additional complexity
Phonological conditioning		
Grammatical conditioning		
Lexical conditioning	Low to high entropy; (depending on the proportion of irregulars/inflection classes in the system of allomorphs and the number of allomorphs)	Low to high additional complexity

Table 2.9: Entropy measures based on the choice of an allomorph given a specific grammatical and phonological context but a non-specific lexical context.

Changing the knowledge of one of those contexts (e.g., knowing the lexeme but not knowing the grammatical category or the set of allomorphs; knowing the lexeme and the grammatical category but not the phonological context) would change the vantage point from where entropy is measured, which would render completely different measures of complexity for different types of conditioning. Nevertheless, there might be a connection to the overall difficulty that lexically conditioned allomorphs pose, when considering which context is more likely to be unknown for speakers. Given the nature of lexemes as an open class, one could imagine that speakers are more often confronted with having to determine the variable lexical than the less variable grammatical or phonological context.

Nevertheless, Table 2.9 shows that entropy is not a useful variable to contextualize the complexity of morphs according to conditioning type; it only shows that lexical conditioning is generally more complex in a system where the paradigm and allomorphs are known, but not the lexeme to which they attach. Because entropy also depends on the size of lexical classes as well as on the size of a phonological and grammatical inventory and the frequency of occurrence, it will not be further discussed here as a potential dimension to typologically operationalize complexity levels of conditioned inflectional morphs. Instead, it will be explored whether the tripartite distinction can apply to user-based complexity, which has the strongest claim to be connected to speaker difficulty.

2.5.3 Complexity based on speaker difficulty

Since the research question of this dissertation claims a connection between morphological complexity and properties of language use, the variables established for this study need to be based on the linguistic reality of speakers and listeners, less so on linguistic description length or entropy measures. The question here is how users react to inflectional morphs that are conditioned by different levels. As mentioned above, description length and entropy might be somehow related to user-based complexity, but evidence is needed to support this claim. However, reactions of speakers to typological comparative concepts might not be as uniform as expected. Typology faces the problem that its cross-linguistic categories often do not align with psycholinguistic effects. This might be due to the fact that typological studies usually work with ready-made grammars where the language is outlined into systemic representations, abstracting away from frequency, prominence, and other properties that affect the choices of users. Even if documentary linguists attempt to be as theory-neutral as possible (such as through a 'basic linguistic theory', see Dixon 2010), grammars usually do not give evidence for which patterns cause more problems to speakers.

There are practical and theoretical challenges in fruitfully connecting both comparative and observational disciplines. A practical challenge arises in bridging the spatial differences between field and experimental linguistics. Typologists owe most of their data to field linguists in remote areas, and it is difficult to 'move the lab into the field' (Norcliffe, Harris & Jaeger 2015: 1023). This explains the lack of cross-linguistic insight between psycholinguistics and typology; experiments are conducted in universities on students of English and other Indo-European languages which makes cross-linguistic generalizations difficult to state. However, the interest and will for a 'cross-linguistic psycholinguistics' is present and growing (see Norcliffe, Harris & Jaeger 2015 for an overview). A theoretical challenge is that typology operates on very abstract and general categories, whereas psycholinguistic experiments must posit a hypothesis containing variables as specific as possible. One single psycholinguistic experiment would likely not be able to determine cross-linguistic degrees of complexity that are needed for this dissertation. Due to the large number of specific psycholinguistic studies, reviews are needed that compile the results under a general theme. As such, trends of specific experiments can be grasped in relation to new research questions, which can motivate typological research.

Until now, there are no studies or reviews that deal specifically with effects of (complexity of) conditioned inflection in general. Instead, there is either research that addresses a specific type of allomorphy, such as effects of the differences between English plural morph variants (Martin and Peperkamp 2020), between grammatically conditioned and unconditioned morphs (Järvikivi, Bertram & Niemi 2006) or between the number of stem allomorphs (Nikolaev et al. 2018). Other studies address difficulty effects resulting from phenomena that are related to the types of conditioning in Section 2.4, like discontinuous forms (Höhle et al. 2006), fusional morphology (e.g. Wagner et al. 2019; Ladányi, Kovács, Gervain 2020), irregular forms (e.g. Prasada and Pinker 1993; Bybee and Newman 1995; Ullman 2001; Pulvermüller, Härle and Hummel 2001; Smolka et al. 2013; Royle et al. 2012, Leminen et al 2019) or inflection classes (e.g. Gross et al. 1998; Rodriguez-Formelis et al 2001, Linares et al. 2006). Besides this research, it seems that effects of morphological complexity are not sought in distinct types of conditioned paradigms but distinct types of morph(eme)s, such as stem vs. affix (Marzi et al. 2018; Balling & Baayen 2012), compounding vs. derivation vs. inflection (Leminen et al. 2019), or simply the number of morpheme categories expressed in a word, such as seen in the acquisition of polysynthetic languages (see Kelly et al. 2014 for a review). Therefore, the effects of complexity of conditioning might not be directly evident as they are overshadowed by effects of other properties.

Typologists must be aware to select the types of morphemes where effects of morphological complexity are clearest. For example, research on processing effort in inflection renders more consistent results than research on derivation, and research on compounding seems to be the least predictive (Leminen et al. 2019). The variability of these effects along this continuum might be based on the variability of different neurological processing systems associated with these distinctions (see Ullman 2001). Although the focus of this study is inflection, the following sections takes into account all types of morphs, but will present differences between inflectional, derivational and lexical morphs if these are mentioned in the literature.

2.5.3.1 Phonological conditioning of morphs

Does the existence of phonologically conditioned morphs (‘morph variants’) contribute to processing cost as compared to the lack of such a conditioning? Phonologically conditioned allomorphs entail a deviation from the one-meaning-one-form principle, and one could imagine that being additionally aware of the phonological context represents a higher cost in processing. Contrary to this suggestion, it has been shown that phonological variants are not necessarily a burden for speakers, because they are ‘naturally’ motivated by their phonological surrounding. In most of the cases, phonological variants emerge as an effect of optimizing production and perception of sounds, by assimilation/dissimilation, cluster–reduction, or preserving syllable boundaries and stress patterns (cf. Nevins 2011). Martin and Peperkamp (2020) show in an artificial language learning experiment on adults that plural allomorph suffixes are better learnt when their vowel is harmonic (some natural features shared) to the preceding vowel (e.g. [pegi-**tɛl**]), as opposed to when their vowel is ‘disharmonic’ (no natural features shared) to the preceding vowel (e.g. [pegi-**tɔl**]). This experiment shows that phonologically conditioned allomorphy alone does not yield uniform results, and that some types of phonological condition can even facilitate the production of allomorphs. Phonologically conditioned allomorph selection not motivated by these optimizing processes exists – like the suppletive morph selection in Kalkatungu ([ku] vs. [ja]) dependent on syllable number –, but their relation to cost/efficiency of phonological or morphological processing has not yet been studied. As such, it cannot be said whether phonological conditioning itself increases efficiency or effort.

Also, there might be differences in processing effort *among* allomorph variants, based on their phonetic properties. Skoruppa, Lambrechts and Peperkamp (2011) show that subjects learn alternations quicker and better when they are motivated by single and not multiple features. For example, a [p]–[v] alternation is more difficult to learn than a [p]–[b] alternation, since the former involves a change in two features (plosive to fricative; voiceless to voiced) whereas the latter only involves a change from voiceless to voiced. In English, children apply Plural allomorphs [s] and [z] more often than [ɪz] to nonsense words (Berko 1958); while the former allomorphs are conditioned by the features [±voiced] of the preceding segment, the latter allomorph combines more features, namely [±strident] [±coronal]. Berko (1958) however suggests that regularity and frequency are more predictive – [s] and [z] occur more frequently in speech than [ɪz]. In contrast, [ɪz] can have a *perceptual* advantage, as has been shown for 3-year-old children with mild-to-severe hearing loss (Koehlinger et al. 2015). Here, subjects produced the allomorph [ɪz] more often than children with normal hearing, indicating an easier perception of the syllabic allomorph in contrast to the monosegmental ones.

These few examples show that phonologically conditioned morphs are not per se more difficult or easier to master during the acquisition process, since their difficulty does not consist in matching form and meaning, but in producing/perceiving phonological properties. Thus, the discussion on the complexity of phonologically conditioned allomorphy requires a discussion on phonological complexity. Phonologically conditioned allomorphy does not seem to be associated with additional effort just because of the nature of its conditioning, and as such, could be grouped together with the variable ‘no allomorphy’.

2.5.3.2 Grammatical conditioning of morphs

While there is no explicit research on whether grammatically conditioned morphs as defined in Section 2.4.2 are more difficult to master than unconditioned or phonologically conditioned morphs, research has found evidence that two parameters associated with this type of conditioning have effects in processing. Morphs might be conditioned by a grammatical category that is expressed elsewhere in the construction; in which case they can appear separated from one another by intervening morphemes. This can be interpreted as co-dependency between morphs; two separated morphs are required to express a semantic content.

In English constructions like *Grandma is always running*, the progressive construction requires the morphemes {is} and {ing} with intervening lexical material. Santelmann & Jusczyk (1998) show that children have difficulties in detecting the relationship of these discontinuous morphemes when they are more than three syllables apart. Similar difficulties in processing co-dependent grammatical units have been found by Höhle et al. (2006) in German perfective constructions, where the choice of the correct auxiliary is usually conditioned by the transitivity of the verb.

(11) German (Höhle et al. 2006: 282)

Pia	hat	ein	Brot	**ge**-kauf-**t**
Pia	AUX.trans.3sg	one	bread	**PERF**-buy(TR)-**PERF**

‘Pia has bought one loaf.’

Here, the recognition capacities of 19-month-olds are affected not only by distance but also by their ability to linguistically analyze the material intervening between the dependent elements. Thus, difficulties arise because of linear distance. However, effects of grammatically conditioned morphs can be present when they are adjacent to their conditioning morph category. In Finnish, some denominal derivational suffixes are invariant, whereas others have allomorphs. These allomorphs are conditioned by specific case/number suffixes. Table 2.10 shows that for *kirjasto,* the denominal suffix *–sto* is invariant across different cases/numbers, whereas for *arvoton*, the denominal suffix *–ton* has other allomorphs (*–ttoma*, *–ttom*).

	–sto	***–ton***
Nominative Singular	kirja-**sto**-Ø	arvo–**ton**–Ø
Genitive Singular	kirja-**sto**-n	arvo–**ttoma**–n
Genitive Plural	kirja-**sto**-jen	arvo–**ttom**–ien
Partitive Singular	kirja-**sto**-a	arvo–**ton**–ta
Partitive Plural	kirja-**sto**-ja	arvo–**ttom**–ia
Essive singular	kirja-**sto**-na	arvo–**ttoma**–na

Table 2.10: Part of the inflectional case-number-paradigm for the Finnish words *kirjiasto* 'library' and *arvoton*, 'worthless' with denominalizers *–sto* and *–ton*. (Cf. Järvikivi, Bertram & Niemi 2006: 401).

Jäkivi et al. (2007) show in five experiments that invariant derivational affixes are more salient (i.e., show shorter response latencies) and enhance morphological decomposition more than the forms with allomorphs – independently of productivity and frequency.

When the conditioning category is not overt, grammatically conditioned morphs can be either analyzed as allomorphs conditioned by a feature of another category, or as cumulative morphs. These forms represent a deviation from the one-meaning-one-form principle and decrease transparency, and are, according to several linguists (Anttila, 1972; Dressler 2005a, b; Goldschneider & DeKeyser 2001; Don 2017; Wagner et al. 2019), more difficult to learn than morphs encoding one meaning (as is the case with agglutinative structures). Wagner et al. (2019) show that cumulative morphs are uttered at a later age than non-cumulative morphs, such as the English 3rd person singular present suffix *–s* (three features), in contrast to progressive *–ing* (one feature) (Wagner et al. 2019: 3050). Non-transparent morphs that are acquired later include grammatically and lexically conditioned morphs, as evidence from several languages with fusional morphology shows (Dressler 2005a). However, agglutinative structures (where morphs encode only one feature) pose their own problems for acquisition, as they exhibit long sequences of suffixes and a variable order of certain morphemes (Dressler 2005a; see also Kelly et al. 2014). Thus, morphs that are non-bound and non-cumulative are generally easier to acquire than bound and cumulative morphs. In sum, grammatically conditioned morphs – whether as allomorphs, discontinuous morphs or cumulative morphs – are not transparent and as such pose difficulties to speakers, in contrast to unconditioned morphs which are transparent. There is, however, no evidence for whether non-transparent morphs pose more difficulty than phonologically conditioned morphs.

2.5.3.3 Lexical conditioning of morphs

The interest in lexically conditioned paradigms, as present also in this thesis, is reflected in numerous studies on the psycho– and neurolinguistic reality of irregular vs. regular forms.

Neurolinguistic event-related potentials like N400, P600, left anterior negativity, etc. have been investigated on the past tense system in English (Prasada and Pinker 1993; Marslen-Wilson and Tyler 1998; Münte et al. 1999; Allen, Badecker & Osterhout 2003, Morris & Holcomb 2005; Newman et al. 2007; Justus et al. 2008, 2009; Kielar and Joanisse 2010; Morris and Stockall 2012; Rastle et al. 2015) and German (Weyerts et al. 1996; Penke et al. 1997; Pulvermüller, Härle and Hummel 2001; Smolka et al. 2013; Smolka & Eulitz 2018; Regel et al. 2017) as well as on inflection classes in Romance languages, such as Italian (Gross et al. 1998), Catalan (Rodriguez–Fornelis et al. 2001) and Spanish (Linares et al. 2006).

These event-related potentials give evidence for distinguishing two different processing systems associated with grammatical vs. lexical elements (Pinker & Prince 1988; Ullman et al. 2005; Ullman 2001a; 2004, Bybee 1995; but see Kapatsinski 2005 for arguments against two-processing mechanisms). Processing of regular grammatical information is operated by a "procedural system" that activates frontal/basal-ganglia structures of the brain, whereas the memorization of elements like lexical items is associated with the "declarative/memory/associative system", which "is subserved largely by medial temporal lobe structures (the hippocampus and related structures)" (Ullman 2001: 46). Declarative knowledge implies that words are retrieved/memorized holistically and are not further decomposed. Procedural knowledge is applied when morphemes of words are processed by analogy to other word forms, which implies that words are further decomposed into smaller entities. While inflectional morphs are more consistently activated by procedural knowledge, lexical morphs are more consistently activated by declarative knowledge. Derivational and compounding strategies can be subserved by one or the other system (Leminen et al. 2019). This has to do with the fact that derivational/compounding morphemes form a stronger connection to the lexical meaning of the stem (Bybee 1985), and as such, lend themselves to being conceptualized together with the stem as one single unit, that means, the unit is retrieved by the declarative system. Inflectional morphemes are less likely to be conceptualized together with stem morphemes since the meaning that they express is more general and less relevant for the lexical meaning of the word. It is therefore easier to generalize differences in effort across inflectional categories.

Regular inflection is in most cases generated combinatorically (Garagnani, Shtyrov, & Pulvermüller 2009), therefore procedurally (Ullman 2004; Whiting, Marslen-Wilson & Shtyrov 2013). Irregular verbs, on the other hand, show evidence for being subserved by whole-word processing (activating the declarative memory network; Ullman et al., 1997; Ullman et al. 2005; Baayen, Dijkstra, & Schreuder, 1997; Marcus et al. 1995; Pinker 1991), but there is *also* evidence that they can be decomposed like regular verbs (Plunkett & Marchman 1993; Stockall & Marantz 2006, Fruchter et al. 2013), in contrast to monomorphemic words/lexical stems or some derivational categories. Thus, irregulars show an ambiguous character as to whether they are decomposed into the lexical base + irregular morph or processed together with the lexical base.

There is no evidence whether the use of either declarative or procedural knowledge (which themselves activate several regions in the brain) is related to difficulty, with exception of older speakers, who rely more on declarative knowledge when learning a second language (Ullman 2001). However, difficulty can be stated for a morphologically complex form, which requires "the parallel activation of the two systems, one of which attempts to compute a form in associative memory [i.e. declarative knowledge], while the other attempts to compute a rule product [i.e. procedural knowledge]" (Ullman 2001: 43; see Pinker and Prince 1991). Would irregulars and not fully productive verb forms be activated by both mechanisms, therefore, require a higher processing effort? This cannot be said for the entirety of irregulars. Irregulars are distinct from regulars by evoking non-straightforward responses. Difficulties may arise when a given form does not directly trigger one or the other system. While frequent irregulars might be retrieved by declarative knowledge alone, a less known form is accessed by declarative and procedural knowledge interacting probabilistically to retrieve it. Failure of the declarative system activates the procedural system, which in some cases results in regularization errors such as *digged* instead of *dug* (cf. Ullman 2001: 44). Regularization, by definition, does not occur with regulars.

Opponents of the dual-mechanism theory (e.g., Justus et al. 2008; Kielar & Joanisse, 2009; Smolka & Eulitz, 2018; Smolka et al. 2013) suggest that regulars and irregulars are computed by the same neural processes, but that irregulars are particularly difficult to process, depending on formal overlap or on frequency. Bybee and Newman (1995) show that chunks that are frequent and familiar are processed by rote [i.e., declarative knowledge] rather than by rule [i.e., procedural knowledge]: very frequent units are subserved by a rote mechanism instead of a rule mechanism, including both regulars and irregulars. Suppletive formations like 'go – went' are stored in declarative knowledge (Bybee & Moder, 1983; Bybee & Slobin 1982). While the dual-mechanism model provides evidence for why irregulars are processed differently than regulars, single-route models highlight the factor of form-meaning overlap/frequency.

How does the picture look in more productive, yet lexically restricted inflectional patterns, like in Romance languages? Research on Catalan (Rodriguez–Fornells et al. 2001) shows that incorrect stem vowels (which express grammatical and lexical information as shown in

Table 1.1 for Spanish) of productive and non-productive inflection classes elicit in the brain a 'left preponderant negativity', which is associated with the procedural system and decomposition. This shows that these stem vowels behave like other inflectional affixes. That means, speakers process lexically conditioned thematic affixes *also* analytically, and these play a role in regularizations. Because children must also learn restrictions to regularizations, there are more errors. Regularization between conjugation classes have been observed in French (Royle 2007) and Spanish children (Fernández-Dobao and Herschensohn 2020). In contrast, Spanish (Clahsen et al. 2002) and Italian (Say and Clahsen 2002; Orsolini et al. 1998) children, rather than regularizing *across* classes, change irregulars (i.e., forms that do not follow entirely the pattern of one class) to 'regulars' (i.e., forms that follow the pattern of the class to which the irregular is classified). The reverse phenomenon – irregularization – is not witnessed for Spanish and French (Fernandez–Dobao and Herschensohn 2020) but can occur in English (Bybee & Slobin 1982). That means, irregular forms that do not bear resemblance to other irregular forms are stored in declarative knowledge and do not invite speakers to generalize their pattern to other forms as inflected forms of semi-productive inflection classes do. Due to their phonological similarity, allomorphs grouped into semi-productive inflection classes might have facilitating effects in that they can be accessed by procedural knowledge and don't need to be always stored in the lexicon like irregulars. However, these 'sub–regular' allomorphs, as standing between fully productive patterns and irregulars, also lead to more errors; an inflectional system containing a clear regular and a clear irregular pattern will only cause errors in the production of irregulars, whereas subregular allomorph systems can be both the source and the target of regularization. This two-fold character can be witnessed for larger groups of similar irregulars in English; for example, instead of *brought*, children utter *brang*, the source of regularization being verbs that form simple past in [æ], like *sang* – but children also utter *sanged* instead of *sang* (Xu and Pinker 1995), in which case verbs from the [æ] group are the target of regularization from verbs that form simple past by *–ed*, like *banged*. However, regularizations of English irregulars are still rare (Taatgen & Dijkstra 2003), because groups of irregulars are small and not as large as inflection classes in Romance languages.

What are the insights from these studies in the evaluation of degree of morphological complexity? First, it seems that irregular morphs and lexical classes represent a challenging analysis for dual-route and single–route approaches; especially semi-productive and semi-frequent lexically conditioned allomorphs appear to activate both procedural and declarative knowledge, or at least, lead to longer latencies in choosing the correct mechanism. These insights fit quite well the intuition promoted in Chapter I that lexically conditioned inflection is especially difficult to master. For largely productive and regular allomorphs, procedural knowledge is activated in first and second language learners, and single irregulars and suppletive forms are mostly activated by declarative knowledge. Furthermore, these semi-productive allomorphs are prone to be the target and source of regularization, as are allomorphs from conjugation classes. This fact might suggest that systems with several 'sub-regular' allomorph sets are more difficult to master than systems with few irregulars/suppletives and one prominent regular class. This runs parallel to what has been concluded with regards to how lexically conditioned allomorphy is complex in relation to description length and entropy, where classes distributed over a similar number of lexemes have a high entropy (Ackerman & Malouf 2013). While these findings apply to inflectional morphology, it is not necessary to look at the effects of lexically conditioned derivational morphs, since the research question specifically targeted the difficulty of inflectional paradigms, and not derivational ones. Nevertheless, this leaves the question open whether grammatically conditioned inflectional morphs behave different from or like derivational morphs.

2.5.4 Conclusion

Section 2.5 investigated whether conditioning of morphs and their paradigms can be used as a variable to show different degrees of complexity, and thus difficulty for speakers. Classifying morphs into types according to their conditioning is a practical way to investigate the cross-linguistic effect of paradigmatic complexity on speakers, and in turn on the structure of a language. In sum, the distinctions 'phonological, grammatical and lexical conditioning' defined on the basis of Anderson's parameter *complexity of allomorphy* are not fully justifiable in terms of description length, entropy and psycholinguistic measures. Nevertheless, there is enough literature to assume that lexically conditioned inflection is linguistically complex and difficult for speakers. This increased difficulty is however not related to the difficulty posed by grammatically or phonologically conditioned morphs. The psycho-/neurolinguistic studies show that there is fewer evidence for a difficulty of grammatically conditioned inflection as opposed to non-grammatically conditioned inflection than for lexically conditioned inflection as opposed to non-lexically conditioned inflection, and even fewer for a difficulty posed by phonologically conditioned inflection as opposed to non-phonologically conditioned inflection.

For phonologically conditioned allomorphs, the difficulties are rather related to articulatory/auditory properties than to the fact that there are several variants of one feature; this would also preclude the question of whether there are meaningful differences between derivational or inflectional morphs that are phonologically conditioned. For grammatically and lexically conditioned allomorphs, difficulties arise from non-transparency, which includes discontinuities and cumulative morphs (although Järvikivi, Bertram & Niemi 2006 explain the higher salience of morphs by the absence of allomorphy); however, this non-transparency does not require a distinction between derivational and inflectional status. For lexically conditioned allomorphs, the difficulty arises when processing systems interact, leading to over-regularizations or hesitations among speakers; however, this pertains specifically to inflectional allomorphs, and single irregulars or suppletive forms are excluded. Derivational allomorphs might not show a clear difference in processing between unconditioned and lexically conditioned types since they are generally more likely to be processed by declarative knowledge alone as they are more relevant to the stem.

Taking the evidence from description-based, entropy and psycholinguistic measures of conditioned morphs, one could say that the most evident cause for complexity may be the activation of a processing system with regards to lexically conditioned inflection vs. non-lexically conditioned inflection. The property of transparency might explain why both grammatically and lexically conditioned paradigms are more difficult to learn than unconditioned morphs, in contrast to unconditioned/phonologically conditioned paradigms; however, there is no evidence for transparency effects with regards to phonologically conditioned paradigms. These two parameters are summarized in Table 2.11.

Allomorph system	Transparency	Processing system[a]	Add. morph. complexity
No conditioning	Transparent	(More likely) procedural	No evidence for additional complexity
Phonological conditioning	No evidence for transparency effects		
Grammatical conditioning	Additional grammatical or lexical information, therefore less transparent		Evidence for additional complexity in one parameter
Lexical conditioning[b]		Procedural and/or declarative	Evidence for additional complexity in two parameters

Table 2.11: Additional complexity associated with conditioning levels of allomorphy across psycholinguistic parameters. [a]Regarding only inflectional allomorphs. [b]Except single irregulars or frequent groups of irregulars.

As can be seen in Table 2.11, there is some evidence to interpret lexically conditioned inflection as most difficult, grammatically conditioned paradigms as less difficult, and phonologically conditioned allomorphy as least difficult for speakers. However, this cline is only induced from two parameters, and difficulty might result from other factors as well. In addition, evidence from description length that states that lexically conditioned allomorphs are more complex than grammatically conditioned allomorphs might strengthen this idea. Entropy, on the other hand, might only be helpful in corroborating the idea that lexically conditioned paradigms are more complex than non-lexically conditioned paradigms; however, this complexity only arises when speakers have to guess a morph not knowing the lexeme that it is attached to.

2.6 Conclusion: more and less complex conditioned morphs

This chapter has investigated Anderson's (2015) parameter *complexity of allomorphy* as a potential strategy to capture and operationalize the difficulty immanent in lexically conditioned inflectional paradigms. The reason to pursue Anderson's (2015) distinction is that it provides discrete set of variables that can be investigated cross-linguistically. The complexity that is purported by this distinction may, to a limited degree, be translated through description length and entropy measures that could numerically justify the operationalization of these conditioning variables. This provides a basis to understand whether complexity in language is related to speaker difficulty. However, description length and entropy also have problems as they are dependent on the size and number of paradigmatic relationships and are dependent on the vantage point from where non-predictability/entropy is measured. Moreover, speaker-based evidence shows that the effects attributed to different conditioning levels are not based on the number, size or simple predictability of a paradigm, but to articulatory properties, the transparency between form and meaning of morphs, and the activation of processing systems. Nevertheless, description length and entropy might indirectly be related to the difficulty posed by opacity and the hesitation to activate a system that better processes lexically conditioned inflection. For example, the higher effort in acquiring cumulative morphs can be explained by a having more than one grammatical category for one form, which increases the description length of a morph. On the other hand, a higher error rate for acquiring/retrieving irregulars can be explained by a lower predictability (higher entropy) faced by speakers and machine learning. Table 2.12 contextualizes the evidence by these different measures, which may reinforce the idea that different conditioning types may indeed increase the complexity and thus difficulty of inflectional paradigms.

Parameter / Conditioning	Entropy[b]	Difficulty (inflection)	Description length	Evidence for increased complexity
None	none	none/no evidence	none	none
Phonological			low to middle	none/low
Grammatical		middle (one parameter)	middle	low/middle
Lexical[a]	low to high	high (two parameters)	middle to high	middle/high

Table 2.12: Evidence for increased complexity from unconditioned, phonologically conditioned, grammatically conditioned and lexically conditioned morphs considering entropy, psycholinguistic difficulty and description length. The different parameters of evidence are not commensurable, unlike the shading of the cells suggests. [a]Except irregulars. [b]From the perspective where paradigm morphs are guessed while not knowing the lexeme.

As one can see in Table 2.12, there is evidence that complexity increases with lexically conditioned morphs in contrast to grammatically conditioned morphs, and also evidence that complexity increases with grammatically conditioned morphs in contrast to phonologically conditioned morphs. These measures are not commensurable, although they may be related. However, the most important distinction is the one that has been obtained from speaker-based, psycholinguistic complexity, since these relate to how inflectional complexity is distributed in utterances.

Taking into consideration Anderson's (2015) parameters as well as their information-theoretical and psycholinguistic dimension, is it plausible to use a tripartite distinction between phonologically, grammatically and lexically conditioned inflection to research the stability of difficult-to-learn patterns? The literature has shown that implementing such a distinction can be plausible, based on the findings according to which lexically conditioned allomorphy is specifically difficult to master. While the other two types, phonologically and grammatically conditioned inflection, show less evidence for consistent difficulty effects, these two categories can serve as control variables to populate the remaining inflectional categories. This also raises the question how to treat unconditioned inflection, the case where paradigms have neither allomorphs, syncretisms nor discontinuous morphs. According to the literature, it might be plausible to cluster unconditioned inflection together with phonologically conditioned inflection, since in this latter category conditioning is not associated with complexity. It is necessary to cover *all* inflectional morphs and paradigms, in order to contrast lexically conditioned inflection with inflection in general, as well as with other types of conditioned morphs. Whereas Anderson's (2015) distinction reduces the parameter where conditioning might correlate with complexity to three variables, it is necessary to expand this distinction beyond the scope of allomorphy and narrow it down to inflectional morphs. The reason is that with inflection, psycho- and neurolinguistic studies render the most consistent results. Moreover, inflectional categories lend themselves better to a cross-linguistic comparison since they are frequent and semantically more consistent, whereas derivational morphology is not always found in the languages of the world, and its semantics are more peculiar or lexicalized. There is also generally less research on derivational complexity (some examples include Nichols 2011a; Stump 2019; Henri et al. 2020). Derivational allomorphy or stem suppletion are less likely to involve the procedural knowledge system, as is also the case with single or smaller groups of irregulars. Instead, allomorphs of larger lexical inflection classes are more likely to show difficulty effects in processing and therefore will likely show complexity effects with other typological variables.

This section has not investigated the effects of multiple conditioning levels, such as when a morph or a paradigm is conditioned by lexemes, grammatical categories and phonological processes at the same time (see Section 2.4.4). It would be too ambitious to reserve a special variable for this phenomenon. To account for this increased complexity, the most complex conditioning type within multiple conditioning could determine the level of complexity that is assigned. For example, an allomorph that is phonologically and grammatically conditioned would be treated as a grammatically conditioned allomorph; an allomorph that is grammatically and lexically conditioned would be treated as a lexically conditioned allomorph. This would increase the instances of complex inflection, which are valuable for this study. This will be further discussed in the Methodology in Chapter IV.

In conclusion, this chapter showed that different types of morphs, defined by their conditioning structure, could be used as variables to show cross-linguistic effects of difficult-to-learn inflection. The practical benefit is that conditioning types can be relatively easy found in grammars, are concrete, make clear reference to universal structures (every language has a phonology, a grammar, a lexicon) and manifest operationalizable degrees of complexity without creating conflicts in assigning values. This allows them to be used as variables for inflectional complexity without measuring the exact description length, entropy or difficulty to speakers in individual languages.

CHAPTER III: FACILITATIVE AND STABLE STRUCTURAL ENVIRONMENTS

3.1 Introduction

In Chapter II, it has been argued that certain types of conditioned paradigms, and lexically conditioned inflection in particular, pose challenges for speakers. Based on this evaluation, one would expect that more complex morphs are more short-lived than less complex morphs, because functional-adaptive pressures in the system would lead to simplification. However, the reality looks quite different: From Indo-European to Latin to Romance languages, conjugation classes have persisted. Lexically conditioned stem alternations may remain in languages for millennia. In fact, complex morphology may be among the oldest 'artifacts' from linguistic history, as Dahl (2004: 264) writes:

> "Some elements of languages, such as (...) phenomena like ablaut in Semitic, are probably among the oldest traceable and still current cultural phenomena. The only serious competitors are general technological innovations such as agriculture. More specific elements of human culture such as religions, calendars etc. seldom go much further back than 2,000–3,000 years."

Furthermore, while there are differences in the distribution of conditioned paradigms, all types can be found across several language families of the world, as will be revealed in Chapter IV. The fact that complex morphology lacks evidence for functional-adaptive constraints, one could ask why it is preserved in the first place. While language-external factors for the stability of complexity have been identified in studies (see Section 1.2), this chapter investigates the possibility of structural effects playing a role in this stability. In order to search for these factors, literature is presented that relates to the likelihood for certain linguistic structures to be acquired and preserved. Two promising properties are surveyed here: prominence and frequency. This chapter surveys literature in order to detect effects that prominence and frequency have on morphology. Prominence and frequency generally facilitate language processing, but – as will be shown – they do not always stabilize structure. Thus, in order to know which structural environments favor the distribution of complex morphology, one has to detect which properties in structure are facilitating, and which structures that have these facilitating structures are stabilizing contexts for morphology. The survey concludes that prominent syllables and obligatory morphological positions can be operationalized as variables to investigate the effects of prominence and frequency on the distribution of complex morphology.

3.2 Functional-adaptive properties and structural stability

In functional linguistics, linguistic phenomena are explained by functional principles – principles that are centered around the purpose in use for interlocutors. If functional principles are not evident for specific constructions, diachronic explanations are preferred that relate changes in structure to the effect of functional principles in the past. For example, it is clear why a language would have interrogative constructions (for the purpose of requesting information from the interlocutor). Likewise, it seems clear that the reason why languages show assimilation of phonetic features in words is the relative articulatory ease for speakers. Other phenomena require more elaborate answers that might not always be agreed upon. For example, one likely explanation for the fact that many languages mark one participant in transitive events (by an ergative or accusative morpheme) is that hearers need to disambiguate between an actor and an undergoer entity in events. However, many languages exhibit no case marking, and in many instances, the roles of actor and undergoer participants are understood from the context. Cristofaro (2019) argues that there is no unifying function in morphosyntactic alignment systems but that they are the product of diachronic developments of constructions that fulfilled a different function in the past. The justification is that one alignment type might arise from sources that are quite different in function, such as ergative affixes developing from instrumental case markers or demonstratives used for topicalization.

In contrast, Haspelmath (2019) argues that these source-oriented explanations as promoted by Cristofaro (2019) cannot explain the trajectory of different pathways leading to the same few constructions (i.e., ergative constructions are common across the languages of the world, even if their source is different). Haspelmath (2019) proposes result-oriented 'functional-adaptive constraints' (p. 7), namely explanations for "how systems come to have properties that facilitate communication." One example of a functional-adaptive constraint is effectivity: "phonological inventories favour five-vowel systems because these make the best use of the acoustic space" (p. 7). For hearers, more distinctive vowels do a better job at distinguishing meaning. Thus, both past and recent phenomena can be explained by these underlying functional principles.

However, a problem arises when trying to explain phenomena like complex morphology, as argued in Chapter I and II. Vestigial traits such as lexically conditioned allomorph paradigms might be too ancient to be traced back by specific development paths, as the quote from Dahl (2004) above suggests (although a source-oriented approach is always possible). Furthermore, these traits lack evidence for functional-adaptive constraints playing a role in the present, as they pose challenges for users (see Chapter II). One common approach is to regard these traits as peculiar or marginal cases that deflect attention away from more frequent, unmarked counterparts. For example, marked singulars are extremely rare in the languages of the world, whereas unmarked singulars aren't (Greenberg 1966, Corbett 2000). Lexically conditioned inflection may be rarer than non-lexically conditioned inflection (this is still in need of investigation), but it is not a marginal phenomenon. With regards to allomorphy, Carstairs McCarthy (2010) states that it is quite pervasive in morphology, and the case where there is no allomorphy is not necessarily more frequent. Furthermore, lexical conjugation classes and irregularity permeates many language families from all continents, like some widely spoken Indo-European and Semitic languages. Viewing these patterns as the result of a residue category in contrast to a purposeful one would not do justice to their vast distribution.

This leads to the question whether principles beyond the ones facilitating communication can be attributed to the distribution and stability of these vestigial patterns. Many arguments have been brought forward that relate to geographical and social factors for the stability and distribution of complex morphology, as mentioned in Chapter I. On the other hand, one could always argue that stability itself is functional, manifesting the principle of conforming to convention (Croft 1999). Conformity to convention allows speakers to establish common grounds, and the stability achieved by this principle would also stabilize the morphology of words, and lastly, morphology that poses challenges for learners.

While these functionalist approaches can explain why complex morphology is present in the first place (conformity to convention, historical results, L1 vs. L2 acquisition, geography), they do not ask if structural contexts play a role in its distribution. In this dissertation, the question of why complex morphology prevails is investigated with regards to the distribution in verbal constructions. Functional, historical and socio-geographical explanations would suffice if complex morphs were randomly distributed across utterances. However, because there is evidence for units smaller than words influencing other units (i.e., nearby sounds causing assimilation, or cumulation emerging from fusion of nearby morphemes), it is necessary to look at how these interactions affect morphological complexity. Structural asymmetries in the language might influence the change and stability of phonemes, syllables and morphemes.

But what are examples of asymmetries causing other asymmetries? Are structural asymmetries merely the reflection of biases that speakers have, i.e., governed by functional-adaptive constraints, or do they autonomously influence structure?

Of course, one could make the argument that structural asymmetries are dependent on functional biases. According to Haspelmath (1999: 199), some of the most common constraints stated in Optimality Theory (Prince and Smolensky 1993) that resonate with functional explanations are recoverability, salience, sonority, distinctiveness, animacy and lexical integrity. Haspelmath (2019: 6) further defines functional-adaptive constraints as "what facilitates communication (including processing) for speakers and hearers". One example is that "phonological inventories favour five-vowel systems because these make the best use of the acoustic space (De Boer 2001), and case systems favour overt ergatives for low-prominence nominals and overt accusatives for high-prominence nominals because of the association between roles and prominence status (Dixon 1994)." (Haspelmath 2019: 7). Thus, competing constraints can explain structural diversity. For example, there is a difference in length of words. Some words are more recoverable than others, and usually these are the ones that are short. Less recoverable words are longer, because longer words have more information which in turn aids in recoverability. Thus, length of words might be interpreted as the result of the functional-adaptive constraint 'recoverability'.

Haspelmath (2019: 14) argues that functional-adaptive constraints can drive language change, by "increasing the probability of change toward a particular kind of outcome, without determining the way in which the change comes about". They are visible in the

> "fit between the causal factor and the observed outcome. If there is a good fit, e.g. if languages overwhelmingly prefer the kinds of word orders that allow easy parsing (Hawkins 2014), or if they tend to show economical coding of grammatical categories (Haspelmath 2008), the best explanation is in functional-adaptive terms, as long as there is a way for languages to acquire these properties." (Haspelmath 2019: 14).

One would think that if a unit has a property that facilitates communication, it would be beneficial for speakers to retain this unit. However, while these properties might influence change, they do not always guarantee the stability of the unit. For example, if a less frequent word is long, this property might facilitate its recoverability. However, the property 'long' does not facilitate recoverability when the word becomes more frequent and therefore more recoverable. As a result, the word shortens and loses the property that facilitated recoverability. This is to show that the goal of a functional-adaptive constraint is not to facilitate the preservation of structure, but communication through structure. Nevertheless, stability of structure is dependent on functional-adaptive constraints. If a structural unit is stable, it means that it has properties that are facilitative for communication, and the forces that change structure are weaker than the forces that stabilize it. From this, an implicational relationship between facilitation and stability follows: all stable units are facilitative units, but not all facilitative units are stable units. This is the basis on which one can investigate which structural environments are more likely to preserve complex morphology. If complex morphology is stable, the structural environments in which it occurs must exhibit properties that facilitate communication; however, only a subset of these environments will exhibit stabilizing properties. Thus, in search for plausible structural environments that predict the distribution of morphological complexity, one has to first narrow down which structural properties are facilitative for communication, and then sort out which of these are stable and may cause stability of other structures.

This chapter investigates two common properties that are facilitative, namely prominence (the contrast and effectiveness of conveying information through prosodic, morphological and semantic structure) and entrenchment (the degree to which a linguistic unit is firmly established in a speaker's mental lexicon, which is dependent, among other things, on frequency). The reason for choosing these properties is that any level of language can have prominent and entrenched units (words, morphemes, phonemes), and it is easier to map prominent and entrenched units to types of conditioned inflection. Other properties might not have such a clear association with functional-adaptive constraints; nevertheless, there could of course be other examples, such as the relationship between recoverability and the length of a unit. Prominence serves an important communicative function in language by allowing speakers to convey information effectively and efficiently to their listeners. Prosodic prominence manifests as a contrast in utterances (Terken and Hermes 2000): An element contrasts with surrounding elements by being prominent. Therefore, prominence contrasts within words allow the mapping of prosodic and morphological structure to uncover distributional associations.

Entrenchment refers to the degree to which a linguistic unit, such as a word or a grammatical structure, is firmly established in a speaker's mental lexicon and therefore resistant to change or variation. Entrenchment can be also characterized as a functional-adaptive property: it helps faster and more accurate comprehension of linguistic units and can be thus seen as facilitating communication. Other than prominence, entrenchment is not constituted by a contrast, but is gradient (Croft and Cruse 2004: 292); very entrenched units can occur next to other very entrenched units. What makes entrenchment useful for analyzing the distribution of morphological complexity is that any type of morpheme can have a variable degree of entrenchment.

However, because entrenchment is a property of the neural network of speakers, its degree in certain parts of words is not directly evident from the structure of these parts. Rather, one has to search for non-cognitive correlates of entrenchment. Frequency of occurrence is a paradigm factor for entrenchment (Langacker 2008; Mukherjee 2005; Schmid 2000; but see Schmid 2010 for reevaluating the idea), and it can be quantitatively measured through corpus studies; more frequent linguistic units are likely to be more entrenched than less frequent ones. Because of the strong correlation between the frequency of linguistic units and their entrenchment, I will examine the correlation between frequency of occurrence and stability of units.

The more important question is whether prominent and frequent units are more stable than non-prominent and less frequent units, and whether this stability affects the stability of the phonological and morphological structure of these units. As mentioned above, not every environment that has functional-adaptive properties is also stabilizing. As such, the discussion of which prominent and frequent units are most likely to cause stability is the main focus of this chapter. In Section 3.3 and 3.4, I show the facilitative effects that different prominent and frequent units have on language structure and investigate whether these effects also lead to structural stability. The facilitative potential of the units is shown in studies that research the ease of acquisition, production, perception and memorization, and the stabilizing potential is induced from diachronic studies that show the rate of their change. Section 3.5 of this chapter concludes the discussion by proposing which prominent and frequent units are most likely to act to stabilize morphology. I then discuss how I will use these units as variables in the cross-linguistic study.

3.3 Prominence

In general, prominent properties facilitate communication by virtue of 'highlighting' some parts of utterances in contrast to others. In the literature, 'prominence' is often used interchangeably with 'salience' (Ellis & Jones 2009, Kohler 2008). Prominence can be defined as follows: (Falk 2014: 19).

> "Prominence […] is a perceptual correlate of structural properties of the linguistic signal. It is linked to the perception or production of a relation between a foregrounded outstanding event and its context, i.e., events differing from their context by means of structural, language-dependent properties."

Prominence is directly correlated to attention-drawing effects in interlocutors. It is a relational property since the prominent element requires a context in which it stands out for the perceiver (Baumann and Cangemi 2020; see also Terken and Hermes 2000). Also, prominence is always linked to a linguistic signal produced in immediate speech events, and as such contrasts with other signals that are part of the speech event. Prominence is tightly interconnected with salience, which Falk (2014: 5) defines as "represent[ing] a cognitive time-dynamic evaluation of discourse-relevant information" (Falk 2014: 5). The increase of salience is especially dependent on syntactic and semantic prominence (Rose 2005). However, prominence is not the only contributor of salience; for example, the "unique meaningfulness of the word ... as a whole" has been identified by Osgood and Hoosain (1974) as a factor contributing to salience of words. As Schmid (2007: 119, emphasis in the original) writes, "irrespective of how a cognitive unit has been activated, it is said to be *salient* if it has been loaded, as it were, into current working memory and has thus become part of a person's center of attention." and furthermore "deeply entrenched cognitive units are more likely to become cognitively salient than less well entrenched ones." Thus, while many factors contribute to salience, prominence and entrenchment are very important ones. In the following, I will use 'prominence' not as a purely cognitive notion, but as a property that pertains to linguistic structure (even if it is dependent on a cognitive interaction). Phonetic, phonological and morpho-syntactic elements can be prominent, and semantic components can be prominent as well – if semantics is understood as conventionalized components of language that contrast with less prominent components.

The effects of different types of prominence can be introduced by an example from Navajo. As shown in Section 2.4.3, every verb stem in Navajo has a lexically conditioned set of inflectional forms. Mastering the stem forms is a highly difficult challenge – especially for L2 speakers – due to the effort in learning different stem sets for every verb. However, despite this complexity it has been shown that the stem is the morpheme that children acquire first, prior to more regular affixes (Chee 2017; Saville-Troike 1996). How can this priority be explained, despite lexical conditioning benefitting communication less than other types of conditioning? According to Chee (2017), the stem syllable is acquired earlier because it contains a bundle of prominent properties that facilitate acquisition for children (Chee 2017: 125-126).

Stress: The verb stem is always monosyllabic and the bearer of primary stress of the word (Hoijer 1945 and Holton 2000; although see Kidder 2008 for the problematic notion of 'stress' in Navajo that she replaces with 'prominence'). As such, the stem syllable stands out in the context of unstressed syllables.

Phonemic contrasts: Stem syllables have the largest number of phonemic contrasts, since many consonants and vowel properties (including contrasts in duration, nasality and tones) can only be found in the stem morpheme (McDonough 2003; see Holton 2000 for other Athabascan languages). The stem stands out by exhibiting phonological features not present in other prefixes.

Position in the utterance: The verb stem is the last morpheme in the word, and since verbs are usually at the end of the sentence, it is also often at the end of utterances. Final positions are prominent because children pay special attention to the ends of units given that they are followed by a pause (Peters 1983; Slobin 1973). A final position stands out in the presence of a following pause.

Semantic content: The stem is semantically prominent in contrast to the meaning of other prefixes "due to the information it contains, the [...] primary meaning [...] needed for a verb construction" (Chee 2017: 126). In this sense, a primary meaning, which is supposed to be more concrete and lexical than the one of other affixes, is more prominent for children. Therefore, the stem stands out because its concrete meaning draws more attention than the meaning of grammatical morphemes.

The properties listed above show that phonological and semantic prominence can explain why children acquire the stem first despite the difficulty of choosing the correct allomorph. Thus, these properties are independent from the form-meaning mismatch inherent in lexical conditioning. This is reflected in children choosing incorrect form-meaning mapping for stems but uttering them in a phonologically conventional way that reflects the prominent features of stems (CV or CVC shapes, longer duration, more phonological contrasts) (Chee 2017). However, the facilitative effect that prominence provides cannot explain why the simpler, yet 'incorrect' stem forms that children utter aren't propagated. One would think that the unpredictable sets of allomorphs would undergo extinction after a couple of generations and that the facilitative effect of prominence might not necessarily influence the longevity of forms. However, diachronic studies in Athabascan have shown that across the entire family, stems and inner prefixes have retained their formal and semantic properties for a relatively long time, including their unpredictable, lexically conditioned quality (Hymes 1956, Thomason 1980, Rice 2000; Mithun 2011; Denk 2019). Those languages in which stems are not always prosodically prominent exhibit less stem alternation, such as Hare and Tetsǫ́nt'ine (Rice 2005; Jaker and Howson 2022). Thus, prominence might indeed influence morphological stability, but the way in which this influence manifests is unclear. As mentioned in Section 3.2, facilitative properties are the precondition for structural stability, but the specific structural environment could also play a role. Thus, one has to investigate prominence effects in relationship to the structural context in which they occur; these structural contexts can differ in size and type. For example, the stabilizing effects might be different with regards to prominent phonemes or prominent morphemes. Also, stability might be dependent on different types of prominence, such as stress, tone or word boundaries. While it is not the purpose of this dissertation to provide a comprehensive survey of all manifestations of prominence and the contexts in which they occur, I will present the prominent units that have been commonly associated with facilitation and stability in the literature. Whether these units can be mapped onto different types of morphs will be also be discussed in the following sub-sections.

3.3.1 Phonological prominence

Phonology can be prominent because segments, syllables or larger phonological structures are prominent. Prominence relations exist between different vowel and consonant qualities, tones, syllables and words. Common phonetic means to mark phonological prominence are a higher F0 frequency and amplitude on the auditory side, and more extreme movement on the articulatory side. However, phonological prominence is also achieved through phonemic contrasts and length. The following subsections discuss the facilitative and stabilizing effects of prominent sound units.

3.3.1.1 Prominent segments

Starting with single segments on the phonetic/phonological level, Baumann and Cangemi (2020: 2) note that the nucleus of a syllable is "more prominent than its onset or its coda." Vowels and consonants are the most central contrast in phonology (Ladefoged 2001; Ladefoged and Disner 2012). Vowels are louder and longer than consonants (Repp 1984), are perceived more clearly by infants (Bertoncini et al. 1988, Bouchon et al. 2015), especially in the womb (Granier-Deferre et al. 2011). This perceptual prominence is paralleled by the fact that languages across different families show more stability in their vowel quality inventories than in their consonant inventories (Cysouw and Dediu 2013: 13–14). However, consonants and vowels also have a different 'division of labour' in acquisition; "consonants are mainly involved in word processing, whereas vowels are favored for extracting and generalizing structural relations" (Hochmann et al. 2011: 1445). This division can be also seen as one that facilitates processing of the lexicon (consonants) vs. grammar (vowels) (Nespor, Peña & Mehler 2003). An extreme, yet clear case of this division of labor is found in Semitic verbs. Roots are associated with the lexicon, whereas vowels with grammatical categories (see Berent, Vaknin and Markus 2007; Prunet, Béland & Idrissi 2000). However, this correlation might be based on the structural peculiarity of certain language families. Nevertheless, vowels are generally more prominent and stable than consonants.

A distinction can be also made with regard to vowel quality: Low vowels are more prominent than mid and high vowels, due to their greater sonority (Kenstowicz 1996; Burquest and Payne 1993: 101). This is paralleled by the process of reduction of high vowels in fast speech and language change, such as in Greek (Dauer 1980), Japanese (Tsuchida and Recasens 1998; Varden 1998) and other languages (Kuznetsova & Verkhodanova 2019; Easterday 2017: 228; Barnes 2006; Flemming 2004). Whether low vowels can stabilize the morphology that they express has not been studied thoroughly; the instability of high vowels might not imply stability of low vowels. On the other hand, Cysouw and Dediu (2013: 13) find that front rounded vowels, despite occurring rarely across the languages of the world, are one of the most stable structural units found in the Word Atlas of Language Structures (WALS). This stability is however not contrasted with other vowel qualities, but other language structures that are not only phonological. Nevertheless, while high vowels are less prominent and more susceptible to reduction, it cannot be said whether the prominence of low vowels also implies stability of low vowels.

With regards to consonants, it has been shown that longer voice-onset times (VOTs) is a prominent property of plosives. Dutch speaking children between 5–6 months show a higher reaction time difference in the contrast between [p^h]– [p] (longer VOT difference) than between [b] – [p] (shorter VOT difference) (Liu and Kager 2015). On the other hand, older children (11-12 months old) show an increased reaction time for the contrast between [b] and [p] and a decreased time difference for the [p^h] – [p] contrast. This shift is explained in that longer VOTs are more prominent for children in the early phase of acquisition, but this property that facilitates acquisition is lost when they adapt to language-specific contrasts (Dutch has a voicing contrast but no aspiration contrast in plosives) (Liu and Kager 2015). Thus, a specific phonemic contrast appears to override a universal phonetic contrast, which means that the universal contrast is not necessarily facilitative for communication. Furthermore, there is not much evidence for aspirated or affricated stops (longer VOT) to be more stable than plain stops (shorter VOT). From Ancient to Modern Greek, aspirated stops became either plain stops or fricatives (see Table 3.2 in Section 3.3.1.3 of this chapter), whereas plain stops remained the same (see Tucker 1969 for a chronology of Greek sound changes). That is, stops with longer VOTs are not necessarily more stable than plain stops, although this cannot be confirmed by looking at only a few languages.

A good example for prominent but unstable consonants are clicks. Ladefoged and Maddieson (1996: 280) argue that click consonants are perceptually salient (i.e., prominent) and their phonetic qualities favor the adoption by nearby language families (such as from Khoisan to Nguni languages). However, Blevins (2004: 196) argues that borrowing of clicks occurred only in a "highly unusual type of bilingualism, with children exposed to Khoisan sound patterns for most of the critical period, with only limited exposure to Bantu." She also remarks that clicks are likely to decay over time and be replaced by non-clicks (p. 195): "[O]lder Nguni speakers produced dental, palato-alveolar and lateral clicks, while subsequent generations neutralized all series to dental clicks, and new dialects of Nguni appeared where these clicks were replaced by various combinations of non-click consonants (Herbert 1986: 28)." Thus, clicks represent a clear case where increased prominence does not lead to increased stability in contrast to non-click consonants. It should be noted that stability of consonants may not imply prominence of consonants. For example, bilabials, nasals and fricatives are frequent classes of consonants (Maddieson 2013a), since they tend to remain in the system when they emerge. While their absence in the languages of the world is rare, it is also stable (Cysouw and Dediu 2013), such as in Australia (no fricatives). Nevertheless, there is almost no research on perceptual and articulatory prominence of bilabials, fricatives and nasals in relation to other consonants; one example being Nam and Polka (2013) who argue that stops are more prominent than fricatives in infant speech. Therefore, one cannot conclude that prominence is directly related to stability in consonants. This however does not contradict the statement above that stability implies facilitation since other facilitative properties might be present in stabilizing these consonant classes.

In sum, vowels are the segments where prominence and stability are more likely connected, unlike in consonants; this relationship can be compared to Dahl's (2004) quote in Section 3.1, according to which ablaut vowels in Semitic are exceedingly stable. This means that mapping morphs to segments to show distributional asymmetries would have different outcomes with regards to consonants vs. vowels. One could of course only compare the distribution of morphs in vocalic environments and see whether low vowels are a factor that contributes to the stability of complex morphology. However, this would reduce the number of possible morphs to be mapped. For this reason, it is more helpful to look at environments that are larger than segments. Section 3.3.1.2 discusses the effects of syllabic prominence.

3.3.1.2 Prominent syllables

Syllables can be more prominent than other syllables, and often this contrast is interpreted as stressed vs. unstressed syllables. According to Gordon and Roettger (2017), stress is "the phonological marking of one or more prominent syllables within the phonological word." The phonetic properties by which stress is realized include a longer duration and a higher fundamental frequency and greater intensity of vowels (Lehiste 1970). Since stress applies to syllables and not segments, Gordon and Roettger (2017: 8) conclude that a longer duration of vowels *and* consonants in the stressed syllable is "the most reliable exponent of stress across [110 studies of 75] languages." In several languages, these prominent properties can be simultaneously present (see Tkachman et al. 2019), or used interchangeably to mark stress, as noted by Easterday (2017) in Lelepa (Lacrampe 2014: 58). Because stress exhibits prominent properties that make syllables stand out from others, it attracts the attention of interlocutors.

Stress is also prominent in the articulatory sense. De Jong (1995) associates English stressed syllables with hyperarticulated jaw movements that contrast with minimal movements associated with unstressed syllables. In English, hyperarticulation coincides with stressed syllables, because it "is not a combined effect of pitch, duration, and stress, but a direct manifestation of stress" (De Jong 1995: 501). In French, L1 speakers also show a tendency for greater jaw movements during prominent syllables (Smith, Erickson and Savariaux 2019; Tabain 2003; Loevenbruck 1999, 2000), although the prominence is perceived differently than in English (Vaissière 2002). Generally, hyperarticulation involves a "longer duration and tighter constrictions for consonants and more open articulations for vowels […] and less articulatory overlap between consonantal and vocalic gestures" (Easterday 2017: 278; see Beckman & Edwards

1994, Fougeron 1999, De Jong, Beckman & Edwards 1993). Whether stress has more consistent articulatory or auditory manifestations across languages remains an open question. However, this overlapping of properties suggests a higher degree of prominence of stressed syllables in contrast to unstressed syllables.

This prominence facilitates the acquisition of syllables. Not only do stressed syllables catch the attention of infants first, but parental speakers of some languages adopt a special *motherese* style towards them through hyperarticulation of stressed environments to further facilitate acquisition (Karzon 1985). There is also evidence for the facilitation of morphological acquisition through stressed syllables. Studies on languages with many morphemes in a word show that the morphemes acquired first are the ones that are expressed by stressed syllables, such as in Navajo (Chee 2017; Saville-Troike 1996), where children first utter the stressed stem. A comparative study on acquisition of Mayan languages (Brown et al. 2013) shows a similar pattern. The children acquiring Mayan start producing the stressed syllable, and the rate at which they learn absolutive agreement depends on whether the absolute suffix of one Mayan language coincides with the stressed syllable; the "interaction with stress seems to provide the verb suffixes with a tremendous boost in acquisition" (p. 299). Similar results have been obtained from the acquisition of Mohawk, where "the child consistently selected the stressed syllable whether it coincided with a portion of the stem or not" (Mithun 1989: 291). Conversely, children tend to omit unstressed syllables in very early acquisition, as has been shown in English (Frumhoff et al 1992; Gleitman and Wanner 1982; Allen and Hawkins 1978; Ingram 1976; Oller and Rydland 1974) and Hebrew (Berman 1977). In Turkish, where stressed and unstressed syllables differ along fewer dimensions, children are less likely to only produce the syllables with primary stress (at the end of the word falling on suffixes), but instead also utter parts of the stem (Aksu-koç and Slobin 1985).

There is diachronic evidence for the higher stability of stressed syllables as opposed to unstressed syllables. Unstressed syllables are more likely associated with lenition and shortening as opposed to stressed syllables (Gordon 2011, Barnes 2006). This is particularly true for languages that have unpredictable stress (Bybee et al. 1998). Stressed vowels, on the other hand, are more likely to preserve phonological distinctions, especially with regards to vowels, and generally information (Altman and Carter 1989). Stressed syllables can still be the triggering context for changes in language structure, such as fortition of consonants, lengthening or diphthongization (Bybee 2015: 46, 62). For example, diphthongs in Spanish have evolved in syllables that were stressed in Latin, as the change from /ˈpopulum/ to /ˈpweβlo/ demonstrates. This led to some irregular forms in inflection. Certain Spanish verbs show stem alternation caused by diphthongization in stressed environments: /ent**e**nˈdemos/ (understand.1pl.PRES) ‘we understand’ vs. /enˈtjendo/ (understand.1sg.PRES) ‘I understand’. Stress therefore does not necessarily preserve the vowel qualities that are stressed but rather vowel contrasts and prevents the nucleus from being reduced and undergoing elision.

On the other hand, vowel reduction is characteristic of unstressed syllables. Vowel reduction often leads to centralized vowels, such as in English, Portuguese and Russian, causing vowel quality neutralization (Barnes 2006, Bybee 2015), but there is also evidence of vowel reduction towards the corners of the vowel space, such as in Luiseño or Belorussian (Harris 2005: 120–121). What all vowel reduction processes have in common is that phonological distinctions and therefore information content or contrast of the speech signal is lost (Harris 2005; see Bucci et al. 2019 for Coratino, a Southern Italian dialect). Vowel elision can be a result of vowel reduction and take several centuries (Kuznetsova and Anderson 2020) or occur abruptly (Kapatsinski 2018; Kuznetsova & Verkhodanova 2019). In conclusion, stress prevents syllable loss by preventing nucleus loss. Thus, one can argue that with regards to syllables, prominence and stability are connected.

Because stress preserves syllables, it can preserve morphology. Diachronically, there is evidence that stress also stabilizes the morphology. Bybee (2015: 35) argues that the loss of unstressed vowels brings about loss of morphology, such as case distinctions in several Indo-European languages. In English, all case distinctions have vanished from nouns (Table 3.1).

	Singular	**Plural**
Nominative/accusative	*scip*	*scip-u*
Genitive	*scip-es*	*scip-a*
Dative	*scip-e*	*scip-um*

Table 3.1: Inflection for Old English *scip* 'ship' (Bybee 2015: 35).

In contrast, morphology falling within stressed syllables is more likely preserved. The number of stem allomorphs in Athabascan is lowest in the languages where stress does not always fall on the stem, such as in Hare (Rice 2005) or in Tetsǫ́nt'ine (Jaker and Howson 2022), whereas in Navajo, where the stem is always stressed, there are large sets of stem alternations: The stem for 'move solid round object' exhibits 39 different form-meaning pairs (Young, Morgan and Midgette 1992: 12). In conclusion, there is a connection between the facilitative and stabilizing effect that stress provides, both with regards to the syllable and the morphology on which stress falls.

3.3.1.3 Boundaries

Phonological boundaries, that is, the end or the beginning of words or utterances, have been associated with a facilitative role in language acquisition (see Echols and Newport 1992 for a review of the literature). Echols and Newport (1992: 206) show that in English, if a syllable is word-final, it is less likely to be omitted in the speech of one-year-olds than when it is non-final; however, stress additionally decreases the likelihood of dropping in both contexts. Stress and phonotactic position interact in a significant way (p. 207), in that "unstressed, nonfinal syllables [are] particularly vulnerable to omission" (p. 212). Furthermore, infants tend to lengthen the word-final syllable (Dial Albin and Echols 1996; see Seifart et al. 2021 for typological evidence), independently of stress or position in the utterance, corroborating the evidence that final syllables are salient to infants when learning their language. In Japanese, where stress is non-existent (although pitch accent in most dialects makes certain syllables more prominent), children first acquire grammatical elements at the end of verbs, which are also frequently sentence-final (Clancy 1985). Position can also affect larger units and facilitate the acquisition of words that are placed at the end of an utterance, especially when the phrase boundary aligns with the word boundary. For example, newly introduced nouns are more likely to be placed at the end of utterances in child-directed speech (Fernald and Mazzie 1991). Prominence is also present in initial phonological boundaries. Word- and syllable-initial consonants are produced with more force and have a longer duration than syllable-final consonants (Keating et al. 2003). Phoon et al. (2014) show that during the acquisition of Malay, children are more proficient at pronouncing consonants at the beginning than at the end of syllables. Thus, syllable-initial consonants are integrated more quickly into children's speech. In sum, elements near boundaries have a privileged status in acquisition.

The story appears more complicated when it comes to associating word boundaries with stability. One phonological process that can lead to loss of consonants is lenition. According to Bybee (2015: 26) lenition is a type of reduction, which is the phenomenon where the "magnitude or duration of a gesture is reduced." Lenition can lead to 'reduction towards zero', where segments are dropped entirely; for example, the sound [p] "has a tendency to weaken and is sometimes deleted entirely". However, the decreased distribution of voiceless bilabials as compared to voiceless coronal and dorsal stops (Maddieson 1984) may have aerodynamic reasons (bilabial plosives are easier to voice than plosives of other places of articulation; Ohala 1983) or areal reasons (gaps in voiceless bilabial plosives are bound to a linguistic area; Maddieson 2013e)

Ségéral and Scheer (2008) show that lenition of consonants is affected by phonotactic position. Intervocalic consonants are the weakest: "spirantisation [i.e., lenition] in Codas supposes the spirantisation of intervocalic stops: cases where stops spirantise in Codas but not intervocalically do not appear to be on record. The reverse of course is not true: spirantisation occurs only intervocalically in many systems" (p. 138). An example for lenition is the English flapping of /t/ and /d/, which only occurs in intervocalic environments. Spanish shows lenition of voiced plosives in non-initial positions, which means intervocalically, (*nube* [ˈnu**β**e]) and in codas (hablar [a**β**.ˈlar]) as compared to *Barcelona* [**b**arθeˈlona] (Harris 1969, Macpherson 1975, Lavoie 2001, cited in Watson 2006).

This means that segments that are internal to the utterance are less stable than at the edges of an utterance. However, word-final syllables are still considered weak positions where lenition and deletion occur, more often than in word-initial positions. This discrepancy does not fully map onto the general property of segments at word-boundaries to be prominent as discussed above. The picture gets even more complicated when considering that word-final codas are not weaker than word-internal codas, and word-initial onsets are not stronger than word-internal onsets (Ségéral and Scheer 2008: 134): “Cases where consonants are strong word-initially but not after codas, or where final codas are weak but their internal peers are not, do not appear to exist.” This is shown in Greek, where stops have undergone lenition in word-initial position, but not word-internally after obstruent codas (Table 3.2).

	After obstruents	After sonorants	Word-initial
Classical Greek	[op^{h}**t^{h}**almos] ‘eye’	[adel**p^{h}**os] ‘brother’	[**p^{h}**ilos]
Modern Greek	[of**t**almos] ‘eye’	[aðel**f**os] ‘brother’	[**f**ilos]
	retention of stop (t^{h} > t)	lenition of stop (p^{h} > f)	

Table 3.2: Retention vs. lenition of syllable-initial stops from Classical Greek to Modern Greek (Ségéral and Scheer 2008: 156; a word-initial example is added here for comparison).

However, it must be noted that ‘strong’ does not imply strong in an articulatory sense. Fougeron and Keating (1997) show that more extreme articulations are still associated with word-initial onsets. Thus, ‘strong’ in the sense of Ségéral and Scheer (2008) can be equated with ‘stable’, even if the authors’ definition includes the synchronic dimension in the sense that lenition operates as a synchronic rule.

Based only on the insights of Ségéral and Scheer (2008), there is a hierarchy between stronger and weaker consonants that cross-cuts the distinction between internal and non-internal, as well as initial and final, and can be summarized as follows:

strongest position *weakest position*

word-internal onsets after obstruent codas >

word-initial onsets >

word-final codas >

word-internal codas >

word-internal intervocalic onsets.

This hierarchy might not be representative for a large variety of languages; however, it shows that the relation between stability of segments and phonotactic position can be more complicated than just the position in the utterance. In sum, the diachronic effects of word- and syllable-boundaries show a discrepancy with regards to the effects that prominence provides in acquisition. While phonological units are generally more prominent at the edges of words than inside words, they are generally more stable word-initially and weaker word-finally, but most stable after obstruent codas and least stable intervocalically.

3.3.1.4 Phonemic contrasts

Prominence contrasts can also be based on phonemic contrasts. Phonemes that are less frequently occurring or are restricted to specific positions might stand out and catch the attention of the hearer. According to Chee (2017: 126; see Holton 2000), stems in Navajo are also prominent because they "have the greatest amount of phonemic contrasts." Some consonants only occur in the onset of stems, and never in the prefix domain (McDonough 2003: 7). Furthermore, the stem is also the syllable that exhibits all tonal contrasts (two level tones and contour tones). This restricted position and therefore prominent status of certain phonemes and tones is present in other Athabascan languages, such as Tanacross (Holton 2000).

Prominence here is established in two ways. First, the existence of many tones in one position leads to many contrasts. This prominence is established paradigmatically by the potential of a tone to occur in a certain position (predictability). Second, the existence of many tones in one position implies the existence of rare tones (if the proportion of some tones are significantly rare). Rare tones can stand out syntagmatically, contrasting with tones that are more common in syllables that surround them. However, phonemic contrasts correlate with other properties. As mentioned above, Navajo exhibits the full range of tonal, quality and length contrasts in the syllable that is always stressed (the stem). In such languages, it cannot be easily determined whether a syllable stands out due to rare vowel qualities or tones. In English, Portuguese and Russian, some phonemic contrasts are neutralized in unstressed syllables (Barnes 2006). In addition, Maddieson 2013b shows that a high number of tone contrasts is also associated with a high number of vowel quality contrasts (Maddieson 2013b). Using phonemic contrasts as a variable for prominence might be difficult and only possible in languages that do not conflate phonemic contrasts with other prominence properties.

Prominence could be established by tonal contrasts in languages that have several tones and lack stress. If a tonal language has stress, stress is correlated with tonemic contrasts (De Lacy 1999; Yip 2001). However, even if stress cannot be determined for some tonal languages, prominence is manifested in a high number of contrasts (Yip 2001). Despite the question whether tonal contrasts are a cause or a result of prosodic prominence (see Zingler 2020 for the latter treatment), it can be argued that the specific number of tones and the syntagmatic pattern in which they occur determines the potential of contrasts.

In a language where high and low tones are equally frequent, neither will stand out. In a language with more tonal contrasts than high and low, the tones that are rarer or show a greater F0-deviation from the other tones catch more attention. This is reflected in the tonal acquisition by Chinese children. It has been shown that the contrast between Tone 1 (the contour with the highest F0) and Tone 3 (the contour with the lowest F0) is learned earlier since these are the most prominent tones (Tsao 2008, 2017). Tone 2 and 4 are learned later as they are not as perceptually contrastive. Furthermore, there is a 'neutral tone' in Chinese, a tone that aligns to the preceding or following tone and is usually found in grammatical morphemes. The neutral tone is a repetition of the preceding or following tone. Syntagmatically, this neutral tone is not considered prominent since it does not stand out from surrounding tones. In studies where Chinese dialects are interpreted as having stress, stress is never associated with syllables that carry the neutral tone (Wang 2015: 8; Chao 1968).

From a diachronic perspective, there is evidence for a preferred retention of tone contrasts that are further apart in their F0 values. Manange, a Sino-Tibetan language from Nepal, has four tone distinctions, but speakers who live in urban areas (like Kathmandu), have only retained a high-low distinction, while the tones with an F0 value in-between have merged with either the low or the high tone (Hildebrandt 2003: 172 ff.). A further development in which maximally contrastive tones are more likely to be preserved is witnessed across Chinese dialects (Yue-Hashimoto 1996), where tone 1 and 3 – the most distinctive ones for children – are also found in Chinese dialects which have only three tones (Chen 2004). There is also evidence for a preferred retention of specific F0 values. Higher tones are more likely than lower tones to be retained. In Bantu languages, high-tone syllables have developed into stressed syllables, thus retaining their prominent property (Goldsmith 1987).

This shows that the number of tonal contrasts, which is a manifestation of prominence, is not necessarily stable; but those tonal contrasts that are most prominent are more stable than contrasts that are less prominent. On the other hand, higher tones which are generally prominent are not necessarily stable, but can evolve into stress, which can stabilize syllables. In addition, there is no evidence for whether certain tonal contrasts or tones might stabilize morphology. Thus, stability of one level (suprasegmental features) does not imply stability of another level (segmental features). Certain tones do not necessarily trigger reduction, although reduced syllables tend to have neutral tones (i.e., tones that do not contrast with surrounding syllables), as in Chinese. Thus, the prominence of tonal contrasts might only be indirectly associated with stability of vowels or morphemes.

3.3.2 Semantic prominence

Besides phonological prominence, semantic prominence can also play a role in facilitating the acquisition and retention of certain morphological environments. When interlocutors produce and perceive utterances, they interact with specific semantic structure. Semantic prominence appears when the semantic content of one element contrasts with the content of other elements in utterances. Since an utterance conveys various meanings, it entails different degrees of prominence in these meanings. Abstracting away from specific lexical meaning differences there are general asymmetries of prominence. I will focus on a common asymmetry, namely concrete vs. relational meanings, but also mention other types of semantic prominence.

3.3.2.1 Concreteness

There is a basic distinction between words that denote entities (mostly nouns) and words that denote relational information (mostly verbs, adjectives, adverbs, auxiliaries). Entities can be more easily foregrounded, placed into focus, unlike relational information. Asymmetric foregrounding of semantic concepts reflects semantic prominence, such as in the ‘figure – ground’ dichotomy which is pervasive in human psychology and language (Langacker 1991). The relative prominence in parts of speech is reflected in children producing nouns earlier than verbs, at least in English (Gentner 2006, Gentner and Boroditsky 2001; Gentner 1982; Huttenlocher 1974) and comprehending them more easily (Gleitman et al. 2005; Kako 2005; Goldin-Meadow, Seligman, & Gelman 1976; Macnamara 1972). A common explanation is that nouns usually refer to concrete entities, and they are represented psychologically earlier than relational words (‘*relational relativity hypothesis*’) (Gentner & Ratterman 1991). Gentner and Boroditsky (2001) and Bornstein et al. (2004) show that nouns predominate in the early speech of Chinese, Korean and Navajo children (Gentner and Boroditsky 2008). However, children show a preference for verb-first acquisition in Mayan languages, such as Tzeltal (Brown 1998). Brown (1998: 748) explains this difference by stating that this “language puts the communicative load into verbs” and “that is what children will learn first.” However, Brown (1998) shows that within verbs, children learn the more concrete morpheme first (the stem), and elaborate affixation later. Furthermore, semantically richer verbs are easier to acquire than semantically bleached verbs, based on the same reason that concrete concepts and objects are more prominent than abstract relations (Gentner & Boroditsky 2001; Tardif 2005). This points towards the fact that stems are usually more semantically prominent than affixes since they express more concrete meanings than affixes. Besides Mayan languages and Navajo, the prominence of stems holds even when the stem cannot be isolated from grammatical elements. Children early utter the consonants of discontinuous stems in Hebrew and use them to acquire other grammatical morphs (Berent and Shimron 1997, Ravid 2003). This is similar to the case in Navajo, where children begin to learn verbs by uttering single stems without its prefixes that are obligatory in adult speech (Chee 2017).

The diachronic effect of concreteness should be analyzed with regards to the stability of form, not its own stability. It is hard to assess how meanings could be stable since they always depend on the context that speakers find themselves in, although there have been approaches to identify the trajectories and rate of semantic change (Traugott 2017, 2012; Hamilton, Leskovec and Jurafsky 2016; Traugott and Dasher 2001; Fortson 2003). It is rather the case that some meanings can cause more stable forms. The popular list introduced by Swadesh (1952) can help determine which meanings are associated with stable forms, which ranks words according to their potential to be cognates in languages, that is, being stable words. Here, it is words with relational meanings that occupy the first places, such as pronouns (I, you, we, that, who) or quantifiers (all, many, one, two). The first concrete entities, 'woman' and 'man', follow on place 16. Place 16–53 are all occupied by entities that are expressed by English nouns ('louse', 'blood', 'bone', 'tail' etc.), afterwards follow verbs (54–71), and from place 74 onwards one can find words that refer to entities, relations or properties, such as color terms. While Cysouw and Dediu (2013: 2) interpret this ranking as a reflection of frequency of use (especially the high ranking of pronouns), the fact that a group of nouns precedes verbs could be taken as a weak evidence that more concrete meanings related with those nouns are more stable than less concrete meanings of verbs and color adjectives.

However, concreteness of meaning does not imply stability of form. Entities referring to 'person' (rank 18) play a relevant role in grammaticalization. For example, the English word 'fellow', referring to an individual person has evolved to a plural marker *–pela* in Tok Pisin. Thus, the concrete meaning of 'person' did not prevent the word from changing its meaning. Words with concrete meanings might undergo grammaticalization and adopt a more general and less concrete meaning. Significant differences in stabilizing behavior between nouns and verbs is unlikely because specific word classes are the result of semantic class and discourse function (see Croft 2001: 86ff.), as such, one cannot generalize stability based on parts of speech.

However, when focusing on stem vs. affixes, one can see differences in stability. Roots are more likely preserved than affixes. According to Beckman (1998), stems are less susceptible to undergoing vowel harmony, are more likely to attract stress and exhibit more phonemic contrasts than affixes (pp. 191–192). When constructions grammaticalize, stems are the ones that survive more often, whereas affixes might eventually drop. For example, the modal auxiliary *can* in English derived from the verb *cunnan* with the meaning 'to know (how).' Prior to grammaticalization, this verb could have all inflectional endings, such as *cunn-e* (singular subjunctive), *cunn-on* (plural indicative), as the following sentences show.

(1) Inflection of *cunnan* in Old English (Bybee 2015: 127-128)

a. Nu **cunne** ge tocnawan heofenes hiw (Ags. Gospel of Matthew xxii)
Now can (subjunctive) you distinguish heaven's hue?

b. Ge dweliaþ and ne **cunnon** halige gewritu. (Ags. Gospel of Matthew xvi)
'You (2pl) are led into error and do not know the holy writing'

Indeed, there are only few cases where stems vanish completely from paradigms, as reported for Ket (*eat*; Vajda 2004: 11), Burushaski (*come*, *make*, *hit*; Berger 1998: 128-129) and Mian (*transfer/take*; Fedden 2011: 271) for only specific, high frequency verbs. One could see the semantic effect on the stability of stem morphs as the case where more specific meanings can become general, but general meanings as expressed by affixes cannot become more general.

Thus, the semantic specificity entailed in stems might 'feed them through' the process of grammaticalization before they become more general and reduced. However, it is doubtful whether the prominence that lexical concreteness provides is the cause of this prolonged stability as compared to affixes. Rather, the higher informational load of stems might be a more plausible reason; and this is another functional-adaptive property that facilitates processing in speakers.

To conclude with regards to prominence due to concreteness, stems are more prominent because they exhibit more concrete meaning than affixes. Roots are also more stable than affixes, but this stability might not be due to stems being more concrete, but due to the advantage that stems have in informational load.

3.3.2.2 Other types of semantic prominence

While there are other aspects in semantics that can form prominence relationships, such as animacy, or focus (pragmatic prominence), these are not elaborated since this dissertation focuses on the morphological structure of verbs. Animacy is associated with nominal categories. Verbs might contain indexation affixes with animate and inanimate distinctions; however, these distinctions are most often relational, signaling actant relationships or degree of involvement in the event (like in inverse or stative-active systems). One could argue that third person pronouns, the only pronouns that can relate to inanimate entities, are often marked by zero because inanimateness causes reduction of forms. However, the more plausible reason for dropping is a higher frequency of third person indexes as compared to first or second person indexes (Bybee 1985, 1995, 2001).

Pragmatic or discursive prominence could be also subsumed under 'semantic prominence'; however, pragmatic prominence does not establish prominence contrasts within the morphology of words. Discursive prominence can be equated to the category of pragmatic focus (Langacker 2008: 66–73, 418; Boye and Harder 2012). In focus constructions, words or constructions can be made relevant, and this prominence helps children acquire and distinguish words. This facilitative potential has been noticed in the acquisition of Hebrew (Veneziano 1988), where mothers' responses are often focalized, with the purpose of making children stay on topic. Focus constructions are effective in acquiring rare word orders (Gourley & Catlin 1978; Grünloh, Lieven & Tomasello 2011). One could say that focused words might stabilize morphosyntax in that they prevent grammaticalization from happening (Boye and Harder 2012). In many languages, pragmatically prominent units correlate with phonological prominent properties, such as length, loudness, higher pitch (Frota 2014; Büring 2010; Gundel 1988; Samek-Lodovici 2005; Selkirk 1995; Szendroï 2003), although this association is less prominent in Spanish than English and weaker still in French (Cole et al. 2019). Thus, the stability of words in pragmatic focus might be also due to phonological prominence, and disentangling this relationship requires an in-depth study. Nevertheless, because pragmatic prominence establishes prominence relations across words, it is difficult to implement it as a variable for investigating the distribution of morphological complexity *within* words.

3.3.3 Morphosyntactic prominence

As mentioned in Section 3.2, any level of language can have prominent units. Besides phonological and semantic prominence there can be morphosyntactic prominence. Because this is a morphological study, morphological prominence should be considered a possible factor in the study of complex morphology, and syntactic prominence should be excluded.[8] However, it is hard to disentangle morphological prominence from either phonological or semantic prominence. For example, Andrews (1989, 1992) derives morphological prominence from whether morphs are formally or semantically distinct from morphs of the same family or paradigm (Andrews 1989, 1992). Beckman (1998) derives increased stability of morphs from the optimality-theoretical concept of 'faithfulness'. A morph is more likely to behave faithfully if it is in a privileged position, that is, when it is less susceptible to alternations on the phonological, semantic and morphological level. Roots are said to be more faithful than affixes, independently from having a more concrete meaning than affixes. However, it might as well be that morphological faithfulness is a result of other stabilizing traits, such as obligatoriness (stems are always obligatory). In any case, 'faithfulness' conflates more general stabilizing properties, but not necessarily prominence, even though faithfulness constraints are ranked higher for prominent morphemes, such as stems.

On the other hand, Giraudo and Dal Maso (2016) give a specific definition of morphological prominence: "psycholinguistic research has progressively focused on purely formal and superficial features of words, drawing researcher's attention away from what morphology really is: systematic mappings between form and meaning" (p. 6). That is, morphological prominence "emerges from relationships between whole word forms and their parts." Thus, salient morphs are not the ones that exhibit prominent sounds or express prominent meaning, but those which are more likely decomposed from their word where they appear, i.e., identified as single

[8] Of course, morphemes that have syntactic function such as indexation morphemes are not excluded in this thesis, since they constitute a large part of inflectional morphology. Rather, this thesis excludes a discussion on prominence related to larger units such as words and phrases.

form-meaning units that are part of words (see Hay 2001 for differences in relative frequency of morphs facilitating this decomposability as well). If this is the strictest definition of morphological prominence, one would expect morphs where the relationship between form and meaning is symmetrical (one-meaning-one-form) to be more prominent, and where it is asymmetrical (syncretic morphs/homophones and allomorphs) to be less prominent. Thus, a conditioned paradigm with allomorphs, syncretic morphs and cumulative morphs would always be less morphologically prominent than unconditioned morphs. This means that the variable 'morphological prominence/transparency' would be dependent on the variable 'complexity of conditioning' and could not be used for this study. Because of the difficulty to define and apply this type of prominence, it will not be implemented in this study.

3.3.4 Conclusion

Section 3.3 presented evidence that prominence is a multifaceted category and permeates all levels of language. Prominence can make units stable, and these units can stabilize other units, showing a connection between functional-adaptive properties of language and the distribution of certain structures. However, prominence does not always produce a stabilizing effect, and sometimes this effect is due to other factors that could be indirectly connected to prominence. Because this study aims to account for factors for the stability of morphological complexity, only types of prominence that signal differences within words are worth investigating. Therefore, prominence relations with units larger or equal than words were not elaborated. Morphological prominence has been briefly discussed but excluded from further implementation because there is little evidence what morphological prominence might consist of beyond a coincidence of semantic and phonological prominent features. The most convincing definition is that morphological prominence is manifested by transparent form-meaning matchings, and since non-transparency is a defining property of complex conditioned inflection, morphological prominence cannot be used as an independent variable in this study.

Prominence comprises facilitating potential with regards to the acquisition of elements that stand under its scope. However, not all prominent units resist change *because of* prominence. This is why it is important to distinguish between facilitative and stabilizing effects, and this shows that specific structure associated with a functional-adaptive property (prominence) needs to be accounted for when explaining structural stability. Table 3.3 summarizes the evidence of stabilizing effects of prominent units.

Prominent unit	Is more facilitative than counterpart	Is more stable than counterpart	Stabilizes morphology	Useful as a variable
Low Vowels	Yes	No evidence	No evidence	No
Prominent Consonants	Yes	No	No	No
Stressed syllables	Yes	Yes	Yes	Yes
Boundaries	Yes	Depends	Depends	Yes
Prominent contrasts	Yes	Yes	No (not directly)	Yes
Semantically concrete morphs	Yes	No (but stems are more stable than affixes)	No	No

Table 3.3: Facilitation, stability and stabilizing potential of prominent units

Table 3.3 is a preliminary illustration of whether prominent units that are facilitative are also stable than their less or non-prominent counterpart. It also shows whether prominent structure also affects the stability of the morphology that maps with it, and which elements might be more promising for defining variables that show cross-linguistic correlations with morphological complexity.

Stressed syllables, boundaries and prominent contrasts are useful variables for further investigation. Stressed syllables and prominent contrasts are more facilitative and more stable than unstressed syllables and less prominent contrasts. Nevertheless, stressed syllables additionally show a stabilizing potential with regards to morphology, whereas for phonemic contrasts, there is no evidence that they cause stability of the morphology that aligns with it. Nevertheless, phonological contrasts might be a symptom of stability and therefore indirectly associated with morphological stability. With regards to boundaries, the stability and stabilizing potential depends on the type of boundary. Generally, word-initial onsets are more stable than word-final codas, and as such it would be most relevant to investigate morphological stability in those contexts.

The association of prominence and stability in segments may be restricted to vowels. While high vowels are susceptible to reduction, it is not clear whether low vowels are more stable. Even using low vowels as a variable and investigating whether those morphemes that align with low vowels are more stable would restrict the investigation to only the vocalic portion of morphs; consonantal morphs would have to be excluded. Syllables, on the other hand, can include different types of morphs. With regards to semantic prominence, it can be said that concrete meanings facilitate acquisition of certain words and stems, but the stability of stems might not be due to their concreteness, but rather due to their more specific content associated with a greater informational load. These latter prominence contrasts might however not be fully applicable to all types of morphemes since semantic concreteness is associated with stems. In conclusion, the most useful variables for this investigation are stressed syllables, prominent contrasts, and, to a certain extent, phonological boundaries.

3.4 Frequency

As noticed during the discussion of prominence, there can be other functional-adaptive properties that affect the stability of structure. In Section 2, entrenchment was introduced as a factor that helps speakers in learning and in recovering structure. It is also a factor in discouraging change or variation. Schmid (2007, 2020) shows that besides prominence (which he calls 'ontological salience'), the stability of structure depends on the degree of *entrenchment* in the mind of interlocutors. Entrenchment is a concept used in cognitive linguistics (most notably in Langacker 2008) which describes a certain degree of automatization in accessing, producing and deriving units. Entrenchment is therefore a gradual property present in the conceptual space of interlocutors (Croft and Cruse 2004: 292). According to Langacker (2008), entrenchment is the source for the emergence of symbolic units like morphemes, words and clauses: "Units emerge via the progressive entrenchment of configurations that recur in a sufficient number of events to be established as cognitive routines" (Langacker 2008: 220). The more entrenched a unit is, the more salient it becomes in cognition, and can be accessed earlier. Entrenchment can also affect higher-level units – schemas –, which are abstractions across different units in language use. The more entrenched a schema is, the stronger it is in the mind of the interlocutors. This schema is then more likely to be generalized over other units. For example, the English past tense suffix *–d* represents the most entrenched schema of past tense, and new words are inflected with this suffix (Langacker 2008: 233-234). A stronger schema is activated more easily by interlocutors, and consequently, predominates in speech. These schemas, again, can act as sources for other constructions, dependent on their relative strength. But how does entrenchment – a cognitive category – relate to language structure? Langacker states that "in principle the degree of entrenchment can be determined empirically. Observed frequency provides one basis for estimating it" (Langacker 2008: 238). Schmid (2010) reevaluates this idea by stating that the connection is not as plausible as one might think, since there are different types of entrenchment interacting with absolute and relative frequency, and the way this connection can be quantified is not straightforward. However, one can assume that contrasts in high and low frequency of units is associated with contrasts in strong and weak entrenchment. Since frequency is a cause of entrenchment, one could say that frequency is a property of the discourse, whereas entrenchment is the structural or cognitive 'imprint'. One can understand entrenchment as a cognitive property in the sense that speakers have the capacity to entrench linguistic units and use it to their advantage; on the other hand, one can understand entrenchment as a structural phenomenon that causes speakers to prioritize certain structures over others by virtue of those being entrenched (and not necessarily functional). In reality, the discursive, neurological/cognitive and the structural are hard to keep apart when approaching language as a complex adaptive system that sees these phases as interconnected and feeding back into one another. Nevertheless, what this dissertation aims to focus on is the structural part of this feedback cycle: where structures that are already entrenched impact the stability of overlapping structure because of entrenchment, and not because of functional reasons that have established the entrenchment in the first place. On the other hand, and as noted in Chapter I, the degree of entrenchment is hard to measure, and frequency must be used as a proxy to assess the structural effects of entrenchment.

In the next sections, it will be shown that frequency strengthens the representation of units in the mind, and frequent units might strengthen the connection to other units. Frequency helps the acquisition, production, perception and memorization of linguistic structure through entrenchment. However, the relationship of frequency and stability is less straightforward. Some frequent elements also get lost *because* of high frequency. This process whereby elements lose phonetic or semantic substance, and eventually disappear is called by Haiman (1994) *habituation*, and Bybee (2001: 11) attributes it to high token frequency, as opposed to high type frequency. Structure that is repeated too often might be replaced with more effective means of transmitting information, in which case very high frequency has no longer explanatory potential for the retention of the element. Todd, Pierrehumbert and Hay (2019) show in a model that high frequency has different effects on elements depending on the perceptual properties of elements. For example, "[i]n changes that act to increase the acoustic ambiguity of the phoneme undergoing change, the model predicts high-frequency words to change at a faster rate than low-frequency words, and vice versa for changes that act to decrease the acoustic ambiguity of the phoneme undergoing change." (p. 15).

Besides the frequency type (token vs. type), one might assume that the structural unit that is frequent also determines the stability of that unit. In other words: does high frequency of a phoneme lead to the preservation of this phoneme in the same way that high frequency of a morpheme does? This question also relates to larger units, such as high frequency of a word and high frequency of a prosodic pattern. In addition, one could also ask how the stability of one frequent unit influences the stability of an overlapping or included unit (such the frequency effect of a word on its morphemes). For example, a common observation is that phonetic loss is caused by high frequency of words (Bybee & Hopper 2001, Bybee and Thompson 1997).

This calls for a consideration of specific structural units themselves influencing the effects of token and type frequency.

In the following sections, I will present findings that reflect a general pattern. The following exemption must be made. ‘Semantic/pragmatic frequency’ as a variable should not be studied here. Semantic frequency, the frequency at which interlocutors activate certain categories, must investigate not only the structural aspect of language, but the interactional and neurological dimension – how speakers and hearers behave during discourse. In this study, it is not possible to investigate this ‘hidden’ frequency because the data that is used in this dissertation stems from reference grammars that lack this information. In other words, the structural phase of semantic entrenchment which would be of interest in this dissertation (namely, semantic categories being propagated because they are entrenched) is hard to keep apart from the cognitive or discursive phase (semantic categories are propagated because they are essential for discourse or universally frequent). Therefore, I will discuss only effects of phonological or morphological units on phonological and morphological structure.

3.4.1 Frequent phonological structure

The following section describes how frequent phonological structure facilitates the acquisition of phonological and morphological structure, and whether this facilitative effect also implies structural stability.

3.4.1.1 Facilitation

Frequency effects from phonology can be experienced in first language acquisition from very early on. Children imitate frequent sound patterns before producing sentences that match the morphology and syntax of adults. Perceptually, children experience and react to sounds as early as in the mother's womb (Partanen et al. 2013; Lecanuet and Schaal 1996). They also show different reactions between the phonology of their mother language and of other languages at birth (Abboub, Nazzi & Gervain 2016; Fifer and Moon 1994); From the perspective of articulation, phonological primacy can be seen in children's babbling as a maturational stage that enables different types of signals (signed or spoken) (Petitto and Marentette 1991). Quite interestingly, babbling is already influenced by the specific phonetics of the first language children acquire; this represents the earliest stage of adapting the articulation to the speech signal (Cheek et al. 2001; de Boysson-Bardies and Vihman 1991; Vigil and Oller 1976). This is only a fraction of evidence for the primacy of sound patterns in early language acquisition, which various linguists seem to agree on, such as Tomasello (2000), Peters (1983, 1997), Pinker (1989) and Slobin (1985). Frequent exposure plays a role in the entrenchment of phonological schemas or templates. These deeply entrenched templates are better predictors of how infant speech differs from language to language than individual phonemes (Vihman and Croft 2007; Vihman 2010). While children make mistakes in the pronunciation of single segments, it is more likely that they produce a word that reflects the phonotactic structure of another word in the language. These templates, not phonemes, are regarded as the "basic units of phonology" by Vihman & Croft (2007: 714). Evidence for this early adoption of templates are word forms produced in a more consistent manner than the target word, as exemplified by English and Estonian utterances (Table 3.4).

	Child form	Adult target	
English (age 1;6)	[byʃ]	brush	
	[dɪʃ]	dish	
	[ɪʃ]	fetch	
	[ɪʃ],	fish	
	[ʊʃ]	vest	
Estonian (age 1;8)	[isː]	isa, issi	'daddy'
	[asː]	kass	'cat'
	[pisː]	piss	'pee'
	[usː]	suss	'slipper'
	[tisː]	tiss	'teat'
	[usː]	uss	'snake'

Table 3.4: Word form produced by child and adult target in English and Estonian (Vihman & Croft 2007: 693, adapted from Waterson 1971 and Kõrgvee 2001)

Table 3.4 shows that in both languages, children begin producing simpler and more consistent forms in relation to adult forms. These consistent templates are the emergent product of the generalization of frequent patterns perceived by babies from adult language and their own babbling. The schematic structure of these forms varies from language to language. For example, English, Estonian and German children produce CVC shapes much earlier than French children (Vihman & Croft 2007: 708), which reflects the non-universal/non-innate character of schemas; however, both German and French acquiring children make use of duration and pitch cues to segment syllable sequences (Abboub and al. 2016). The templates are, as noted above, product of a higher entrenchment of patterns in the mind of interlocutors, which in turn correlates with their relative frequency of occurrence. For example, rhythm (a predictable pattern of alternation between unstressed and stressed syllables) is acquired earlier in French than in English (Grabe, Post and Watson 1999), reflecting the higher type frequency of a limited repertoire of stress patterns in French. An increased repetition of phonological patterns leads to speakers uttering phonotactic structures before mastering the morphology. This is most evident in language where morphology is complex, as observed by Chee (2017) in Navajo. Here, children's early words include not only bare verb stems, but most frequently two-syllable sequences with one prefix (Chee 2017; Saville-Troike 1996). The syllable does not always match the meaning intended by the children, but contains information from prefixes closer to the stem, which express tense, mood, aspect and subject agreement. These inner prefixes are less prominent than the stem, yet children still utter them before the outer prefixes that are arguably more prominent than the inner prefixes. This is shown in the following examples provided in Chee (2017):

(2) Utterances from a 4.7–4.11-year-old child. The syllable reflecting tense/mood/aspect/subject marked by inner prefixes is underlined.

Attempt	Target
ditąsh	ádiłtąsh (p. 369)
híiní'ą́	iih yíní'ą́ (p. 364)
diish'aah	bá dísh'aah (p. 368)
díínáʼ	nídii'na' (p. 370)
zigai	neezgai (p. 364)
deezdá	dah neezdá (p. 370)

Although children also produce combinations of outer prefixes + stem, it is remarkable that the production of inner prefixes is more phonologically robust. However, this is less surprising when considering that a minimal verb form in Navajo consists of at least two syllables, where the last syllable always aligns with the stem morpheme and the preceding with tense/mood/aspect prefixes. Thus, bisyllabic words are a strong, entrenched schema in the mind of Navajo speakers. Some Athabascan languages show even stronger phonological consistency. According to De Reuse (2005), the template of Apache verb forms can be better defined in phonological than morphological terms. Words have more phonological restrictions than sentences (because morpheme order is more fixed than word order), and in languages with longer words, the phonological restrictions that they inherit improves the development of phonological templates. Thus, entrenched phonological templates, which are instantiated by frequent and consistent rhythm and syllables, can facilitate the acquisition of both the phonology and morphology that they encompass.

3.4.1.2 Stability

With regards to the diachronic stability of highly frequent phonological units, Bybee (2001) argues that phonological change does not happen randomly, but always within a specific context. It is rather the relative frequency of *words* that initiate phonetic change and loss. Mere phoneme frequency does not seem to influence the retention or loss of phonetic material. When contrasts are neutralized over the course of time, it increases the frequency of the phoneme resulting in the neutralization. However, once a phoneme has increased in frequency, it may change its frequency. To take an example from English: the phoneme [ə] is the result of neutralized vowels in unstressed positions. This has made [ə] the most frequent phoneme of the English language (Hayden 1950). The development of English [ə] and its likelihood of deletion (Song 2013) shows that phonemes can neutralize or drop without the language gaining new phonemes, and that high frequency of different sounds might lead to the emergence of a new phoneme, as happened with unstressed lax vowels in English. This is reflected by the unbalanced distribution of phoneme inventories across languages and areas in the world (see Maddieson 2013 c, d). Languages can maintain inventories of less than a dozen phonemes, as in Pirahã (Everett 1986) or Rotokas (Robinson 2011: 25–27). In these languages the frequency of individual phonemes is higher than in a language with a larger phoneme inventory. Because assimilation and lenition are a more likely change than dissimilation and fortition (Millar and Trask 2015: 49–51; Bybee 2015: 46; Mowrey and Pagliuca 1995), some phonemes might increase in frequency and as a result decrease the phonological inventory, which means that the number of phonemes in the inventory is not stabilized due to heightened frequency of individual phonemes. Blevins (2004) argues that specific phonemes or phoneme combinations are not per se resistant to change, and that there is no intrinsic motivation in phonology to change or for a language to have certain phonemes:

> "Sound change happens, but it does not occur in order to make speech easier to articulate, easier to perceive or easier to transmit; it does not necessarily result in a more symmetrical, more stable or generally improved phonological system; for every case where it happens, there is a parallel case where it does not happen." (Blevins 2004: 45).

However, if frequency is associated with larger and more schematic phonological structures instead of single phonemes, then the effects are more predictable. The more entrenched – that means, more frequent – phonological templates of words or syllables will have a stronger effect on which phonemes drop or are retained. Blevins (2009) shows that certain syllables have remained relatively stable in Central Pacific Oceanic (such as Polynesian languages), despite being unstressed. Following the findings on Section 3.3, one could expect that over time, unstressed syllables would lose segments or drop entirely if they are not stressed. The reason why this did not happen in Central Pacific Oceanic languages is that open syllables are so entrenched that it prevents the loss of the nucleus vowel which would result in consonant clusters. Blevins (2009) suggests that deeply entrenched phonological templates like the ones in Central Pacific Oceanic not only preserve these structures, but initiate changes that align with the template. For example, metathesis always "result[s] in consonant clusters which are characteristic of the pre-sound change stage of the language" (p. 35). The tendency of sound change being governed by more abstract phonological schemas than the features of single phonemes reflects the early stages of phonological acquisition in infants as mentioned above. Thus, phonological templates influence the speech in acquisition and in continuing replications of adult speech. On the other hand, speakers of related Western Malayo-Polynesian languages where closed syllables are attested show a higher tendency for dropping unstressed vowels (cf. Blevins 2009: 39).

The same effect can be seen for compensatory lengthening of vowels. Compensatory lengthening is more dominant in languages where there is a clear distinction between short and long vowels (Kavitskaya 2002). This shows that phonemes within structures that are very consistent and frequent might undergo change dependent on the stability of larger phonological structures. Frequent phonological templates not only help children acquire language better, but are stable over time, and can preserve phonemes and morphemes that align with them and change those that do not. Frequent templates can be characterized as partaking in negative feedback cycles: the more frequent they are, the least prone they are to change.

Can strongly entrenched phonological schemas also stabilize morphology? Whether these schemas preserve morphology depends on whether the morphology aligns with them. For example, the emergence of utterance-final stress placement in French has resulted in the dropping of final unstressed syllables. Final unstressed syllables that carried inflectional information in Latin have dropped, so that some French words do not inflect for gender anymore. An example is the merge of Latin feminine and masculine forms, such as /ˈultimus/ and /ˈultima/ ('last') to one form in French (e.g. /ylˈtim/ 'last'). Here, not only has stress shifted to the last syllable of the word, but this last syllable was not the last syllable in Latin, which marked the distinction between a masculine and feminine form. Morphemes that align with a strong phonotactic schema are more likely retained. In Navajo and other Athabascan languages, the bisyllabic requirement for verbs can explain why older morphology has remained stable. Parallel to the early acquisition of pre-stem syllables, it has been shown that the inner morphemes, which are more frequently expressed by the pre-stem syllable, are the most stable across Athabascan languages, in function and order, whereas the outer prefixes have undergone more change (Rice 2000, Mithun 2011, Denk 2019).

In conclusion, highly frequent templatic phonology such as syllables or rhythmic patterns can preserve phonemes and morphemes if their features or segments align with them. Because these templates are reinforced across replication events due to their high frequency (through a stabilizing feedback loop), they not only help the acquisition of morphology but stabilize it.

3.4.2 Frequent morphosyntactic structure

According to Langacker (2008), morphemes, words and sentences represent symbolic structure; however, "a morpheme is definable as an expression whose symbolic complexity is zero i.e., it is not at all analyzable into smaller symbolic components" (p. 16). Highly frequent symbolic (i.e., morphosyntactic) units have various effects on units of other kind, such as constructions, words and morphemes. While high frequency generally facilitates the acquisition and memorization of such forms, the diachronic effects depends on whether these forms have a high type or token frequency, as is explained in the following.

3.4.2.1 Facilitation

In general, more frequent morphosyntactic patterns, small or large, are more likely to be entrenched and therefore better recalled and produced than less frequent patterns. According to Bybee (2006, 1985), high frequency of morphemes and morpheme combinations (words) increases the ease of access. However, parallel to phonology, children do not learn language morpheme by morpheme but instead through morphosyntactic schemas which are sometimes simpler than the ones in the mind of adults. Before mastering symbolic structure, children exploit both semantic and phonetic schemas in which morphemes, even frequent ones, might be undefined. This is especially the case in languages that have complex morphology, such as Navajo (Saville Troike 1996; McDonough 2003; Chee 2017); Quechua (Courtney and Saville-Troike 2002), Mohawk (Mithun 1989) or Mayan languages (Peters 1997, 2001). There is a general strategy for children to use 'filler syllables' that correspond to frequent morphemes in the adult target, elements which have been termed 'proto-morphemes' by Peters (1997, 2001). These proto-morphemes align first with phonological templates found in the language rather than the morphemes of the adult language. The Navajo examples in (2) could be also interpreted as results of morphological frequency rather than phonological frequency. If children only relied on phonological schemas, they would more often produce the most frequent consonants, even if they don't match the adult target. For example, when children pronounce ***di**-tąsh* instead of ***ádi**-ł-tąsh*, 'someone is prissy' (Chee, personal communication) it shows that the morpheme ***ádi-*** 'reflexive' has been acquired, even if partially, and incorporated into the frequent bisyllabic template. Alternative insertions that reflect entrenched phonological schemas in Navajo could be also /*ni*/ or /*yi*/, which are very frequent syllables. According to Chee (2017: 260), children used *yi-* as filler syllables that are phonologically plausible but at times ungrammatical; however, these instances are few. The Navajo examples therefore not only show the strength of the entrenched, frequently recurring phonological schema in the form of consistent bisyllabic verb forms, but also the effect of morphological frequency; some morphemes like ***ádi***- in ***ádi**-łtąsh*, even if they are hard to learn for L2 learners, are produced almost correctly from very early on.[9]

[9] There might be two different interpretations of *di-* in the attempt form ***di**tąsh*. One could relate to hand movement such as *'a-di-ł-tąsh* 's/he flicks'. The other meaning is 'reflexive', as part of the morpheme *ádi-*, as suggested by the target form. Chee (2017) suggests that *di-* is part of the morpheme *ádi-*; in this case, the child has selected the correct morpheme, but did not produce the morpheme with the preceding vowel *á-*.

3.4.2.2 Stability

Highly frequent morphosyntactic structure can also impact the stability of phonotactic, morphological and semantic structure. Bybee (2001) shows that the clearest effects of frequency are seen on the word level. A word repeated very often has a two-fold effect on the retention of language structure. First, "repetition leads to reduction of form", that is, of phonetic/phonological elements. Second, "phonetic change often progresses more quickly in items with high token frequency" (Bybee 2001: 11). Due to entrenchment caused by high frequency, articulatory movements become more automated and reduced (Boyland 1996). On the other hand, high *type* frequency "encourages phonetic change, but it renders items more conservative in the fact of grammatical change or analogical change based on the analysis of other forms" (Bybee 2001: 12; see Phillips 2001). If a morpheme has a high type frequency, this strengthens the paradigm in which it occurs, and as such, the morphosyntactic template. A paradigm itself might also have a preservative effect. Bisang (2014) shows that languages that index subject through person and number features in paradigms will more likely retain these morphs than languages in which the subject is indexed by morphs through other features and do not pertain to paradigms, such as languages from Southeast Asia. This preserving effect of type frequency also applies to syntactic patterns (see Bybee and Thompson 2000, 1997). High token frequency of words might still preserve irregular morphology better than low token frequency words. For example, Bybee (2006: 715) shows that the form *wept*, while having a low token frequency, has been regularized to a form containing the high type frequency suffix *–ed*, whereas the high token frequency form *kept* has remained stable. However, the stabilizing effect is not lasting if the morphological strategy has a low type frequency. For example, the forms of the verb 'to be' (*am, are, is*) have a very high token frequency, yet their alternation pattern only occurs with this verb. As such, it is more likely for these forms to disappear (such as in African American Vernacular English where the forms *be* and *is* fulfill more functions). On the other hand, one could also argue that the forms *am, are, is* are still quite stable (as evidenced in other varieties of English) since they have a high type frequency of occurring in syntactic constructions. Thus, one type frequency relates to the morphological alternation pattern, the other to the syntactic construction in which it occurs. Vowel alternation patterns with a higher morphological type frequency such as [ɪ – æ –ʌ] in *sing – sang – sung* and *ring – rang – rung* are quite stable. Furthermore, morphemes that have a high token frequency *and* high type frequency tend to be stable as well. An example would be pronouns which are one of the most morphologically stable categories, more so than noun phrases (Bybee 2001: 12; Givón 1979); they occur very often in discourse and appear in many syntactic constructions. Affixes that have a high frequency are also less likely to fuse with the stem that they are attached to and resist analogical extension by other, less frequent morphemes. Hoekstra and Versloot (2019) show that in Frisian lexically conditioned plural suffixes have been preserved because of the high type frequency of these plural markers (in addition to prominence). However, Hay (2001) shows that it is rather frequency *contrasts* among morphemes in words (relative frequency) that predicts whether fusion happens. For example, the morphemes in *defended* (defend–ed) differ in frequency: *–ed* has a high token and type frequency, whereas *defend* has a relatively low type and token frequency. As Schmid (2007), Hay (2001) and Geeraerts et al. (1994) argue, stronger entrenchment might not depend on pure frequency, but on frequency contrasts, differences in relative frequencies between morphemes. Lastly, if semantic shift of the word happens, this can also lead to the vanishing of morpheme boundaries, despite the morphemes exhibiting a high token and type frequency. For example, the utterance ‘how are you’ contains words with high frequency, but one lineage of this utterance evolved to the greeting 'howdy' or 'hi', leading to reduction and erasure of morpheme boundaries. In conclusion, while both token and type frequency of words and morphemes lead to phonological erosion, morphology is stable when it has a high type frequency, and unstable when it has a low type frequency; high token frequency of words might facilitate and preserve irregular forms, but it might lead to erosion of morpheme boundaries inside words which in turn might no longer stabilize morphological complexity.

3.4.3 Conclusion

High frequency can have a strong impact on the facilitation and stability of phonology and morphology. With regards to phonological frequency, it is high templatic frequency that causes stability of phonemes and morphemes (and the template itself). With regards to morphological frequency, it is morphemes with high type frequency that tend to remain stable. The distinction between type and token frequency was not implemented with regards to phonological frequency because it is hard to define phonological templates. However, highly frequent phonological schemas could be interpreted as a manifestation of high type frequency of phonemes – phonemes or syllable types that occur in many types of words or utterances might lead to the emergence of strong phonological schemas. In this case, one could simply say that type frequency leads to preservation regardless of the units (phonological, morpho-syntactic) that exhibits it. On the other hand, one could regard morphological high type frequency as a manifestation of high frequency of symbolic (morphosyntactic) schemas. Morphological schemas are morpheme combinations that tend to occur very frequently, and as such increase the entrenchment of this morphological schema. According to this view, one could say that deep entrenchment of abstract templates or 'schemas' leads to preservation regardless of the type of template (phonological, morpho-syntactic). In addition, contrasts in frequency between certain morphs (i.e., stem vs. affixes) might also play a role in the stability of morphs. For the time being, I will conclude that high phonological schematic frequency and high morphological type frequency are the properties associated with stability of phonological and morphological structure. This stabilizing effect might be overridden by other changes, such as when semantic change leads to erasure of morpheme boundaries despite these morphemes having a high type frequency. The following table summarizes the stabilizing effects of phonological and morphological structure.

Highly frequent unit	**Stabilizes...**	
	Phonological structure	*Morphological structure*
Phonemes	No evidence	No evidence
Phonological templates	Yes	Yes
Words and morphemes as tokens	No	No
Words and morphemes as types	No	Yes

Table 3.5: Stabilizing effects of frequency across phonological and morphological units.

3.5 Conclusion and operationalization potential

This chapter discussed some possible structural environments in which morphology remains stable. The takeaway from this chapter is that facilitative properties are not sufficient to deduce stability of language structure. Thus, prominence and frequency are beneficial for the acquisition, memorization and processing of language structure and as such improve communication.

The question of stability is only indirectly related to these functional-adaptive properties and depends more on the interplay of types of prominence and frequency with different structural units. This provides an understanding why complex morphology might survive in the languages of the world. And this survival affects the synchronic distribution of inflectional complexity in specific structural environments.

However, in order to show the relationship between complex inflection and structural environments, it is not only important to identify these environments but to operationalize them in a manner that facilitates a typological comparison. Chapter II proposed a distinction between phonological, morphological and lexical conditioning for approximating the degree of morphological complexity in typological research. Likewise, a parameter is needed that contrasts more and less stable structural environments. The most promising structural types that stabilize morphology are stressed syllables, frequent phonological templates and morphemes with a high type frequency.

3.5.1 Prominence

The most promising variable that relates to prominence contrasts is stressed and unstressed syllables. They can be easily mapped onto morphological structure since stressed syllables might or might not align with certain morphemes. But this variable cannot be utilized for languages that do not have stress or when morphemes are consonantal. To construct a variable based on prominence for these cases, one could have recourse to other prominent units:

a) Low vowels

b) Phonological boundaries

c) Phonemic contrasts

d) Concrete meanings (lexical morphemes)

Low vowels might be stable across time; however, this property is only useful for comparing the distribution of morphemes that are vocalic. While consonants at phonological boundaries might be more stable, the specific segment at a boundary is more relevant than the mere contrast between a boundary and a non-boundary position. This prominent property might only be used for morphemes that consist of consonants alone. With regards to phonemic contrasts, the survey has shown that the quantity of contrasts might not directly preserve the morphology, but that large number of contrasts is an indicator of stable environments. This property could be used as a variable for when a language does not exhibit stress. Finally, concrete morphs such as stems might be more stable than affixes, but prominence in meaning itself does not cause stability of morphology. Rather, the stability of stems might be due to other necessary functions of stems, such as specifying the meaning of the word. This property might not be useful in the mapping to complex morphology, because it is expected that affixes are generally less semantically prominent than stems. However, it might be expected that lexically conditioned stem alternation is tolerated because of the stability of stems, and thus, this variable must be controlled for. In conclusion, stress should be prioritized as a variable of prominence, phonemic contrasts only be used as a variable for languages that do not have stress; boundaries could be a variable for morphs that do not exhibit vocalic segments. The specific implementation of variables related to prominence will be elaborated in Chapter IV.

3.5.2 Frequency

It might first appear difficult to implement frequency as a variable in this typological investigation, given that the data used for this dissertation is taken from grammars and not from corpora. Frequency counts are almost always lacking in grammatical descriptions. However, a solution based on qualitative distinctions could circumvent this obstacle. Inflectional categories are already very frequent in the discourse, especially in verbal constructions. To avoid counting type or token frequency of inflectional morphemes, one could simply draw the line between the morphemes of a category that is obligatorily expressed (part of every verbal construction) and the ones that are not obligatory (not part of every verbal construction). Thus, if a morphological position such as subject indexation or tense is always present, its morphs expressing the features of this category will very likely have a high type frequency. However, obligatoriness of a position would not account for a paradigm that has zero morphs; one would have to account for the type frequency of zero morphs even if they do not obligatorily fill the position. Zero morphs such as third person singular might be the result of phonological erosion, and this can be due to token *and* type frequency. Thus, the concept of obligatoriness must be expanded to include positions that have zero morphs, since these are usually a result of high frequency, and indicates that other morphs in the same position might have a high type frequency.

It is also difficult to account for frequent phonological templates, especially for languages that do not have a predictable phonotactic structure (such as restricted syllable shapes). Nevertheless, these templates are stable because of their high frequency, increasing their entrenchment in the mind of speakers. Phonological templates can also stabilize the morphology if the morphology aligns with that template. However, contrasting the variable 'frequent phonological templates' with complex paradigms would require the analysis of morpho-phonological alignment. This is not easy to achieve, since phonological templates do not necessarily have one anchoring unit (e.g., does a phonological template emerge around word edges or stressed syllables; which is the most frequently occurring part, and which is the less frequent?). One possibility to contrast morphemes with frequent phonological templates is to align them according to pivotal structures such as syllables, feet, or intonation units. But then one would also have to determine what patterns are not part of frequent templates. Some languages exhibit a distinctive contrast between frequently and less frequently occurring patterns of morpho-phonological mapping, like Navajo, where verbs must have at least two syllables and these syllables always align with stem and subject/TAM prefixes. This bi-syllabic template might be more entrenched and therefore more stable than tri- or multisyllabic templates since trisyllabic templates contain bisyllabic ones. This means that in languages in which morphology aligns with an obligatory syllabic requirement, the morphology will be stable.

However, because of the complicated nature of determining phonological 'obligatoriness', morphological obligatoriness should be preferred in operationalizing high frequency of elements within words. Phonological obligatoriness might be chosen exceptionally in cases when it is not clear whether a morphological paradigm is obligatory or not. This could be done by looking at whether morpheme paradigms align with a morpho-phonological position in the word. For English verbs, this means that tense can be marked either on the stem or with a suffix. While the stem is always part of every verb, the suffixes *-(ə)d*/*-t* (past) and *-s* (present) are restricted to certain verbs and to a specific person/number category. The semantic side of the paradigm can be considered obligatory (every verb has a tense), but the phonological side is variable: there is not always a tense suffix (/d/, /t/ or an unstressed rhyme /əd/) in coda position, such as in the verb 'flew' /flu/. However, this does not mean that evidence for zero-morphemes in one paradigm disqualifies the paradigm from being obligatory in the strict sense. Since entrenchment is a matter of degree, those paradigm positions that exhibit morphs for all features could be regarded as both morphologically and phonologically obligatory, since a position is always filled by phonological segments. Paradigms with few zero morphs might show less morpho-phonological entrenchment than morpho-semantic entrenchment. The variable of 'obligatoriness' will be further elaborated in Chapter IV.

In sum, the search for variables that comprise stabilizing structures has revealed that stressed syllables and obligatory morphemes are the first choice for operationalization, and these variables should be aligned with different types of conditioned morphs. Thus, it is expected that more complex morphemes are more likely expressed by stressed syllables and part of positions exhibiting obligatory paradigms.

CHAPTER IV: METHODOLOGY

4.1 Introduction

This chapter presents the materials and methods used to answer the research question whether the distribution of complex inflection is partially determined by structural environments that are stable and are thus a contributing cause for the survival of complex inflection in these environments. While Chapters II and III discussed the theoretical potential to establish typological variables based on usage-based and experimental concepts, the goal of this chapter is to further concretize these variables so that they can be applied to cross-linguistic verbal constructions. The purpose is to create a guide for how to annotate degrees of morphological complexity and structural properties that are associated with the preservation of this complexity. While this dissertation uses a specific language sample to investigate these variables, the annotation guide is applicable to any language, and that the results from this study are reproducible. This chapter also explains which languages were gathered and how the different samples were defined with the purpose of answering the research question.

Section 4.2 discusses biases that are usually present when trying to construct a language sample (genealogical, areal, bibliographic, sociolinguistic, structural). To minimize these biases, genealogically and geographically distant languages have been included, which differ in number of speakers and their morphosyntactic structure. Almost all grammatical descriptions used for this study stem from different authors. The languages are grouped into a main sample (30 languages) and two control samples (6 and 5 languages each). Section 4.3 presents the language sample and discusses to what extent biases have been minimized for this study.

The remaining sections discuss the methods applied in operationalizing the morphological diversity of the languages for comparative purposes. In Section 4.4, it is argued that inflectional categories are the best fit to investigate the correlation between morphological complexity and stabilizing structural units. Inflectional categories are defined in order to target highly frequent and semantically general morphs. Furthermore, it is explained how the morphological structure of verbs has been arranged to serve the purpose of annotation in the spreadsheet. The spreadsheet groups morphs into templatic positions that are then annotated for values related to morphological complexity, prominence and obligatoriness. In Section 4.5, I discuss the operationalization of the parameter 'inflectional complexity' based on the insights in Chapter II. The different types of conditioning are assigned three values (lexically conditioned, grammatically conditioned, unconditioned) which can be annotated for each inflectional position. The operationalization of the stabilizing structural variables as presented in Chapter III will follow in Section 4.6 of this chapter. The different types of prosodic prominence (stress, tone variability, prosodic boundaries) are assessed for their potential to construct these variables. Finally, a tripartite value distinction between always prominent, never prominent and sometimes prominent is defined for the annotation. For obligatoriness, a tripartite distinction into morphologically obligatory, morphologically + phonologically obligatory, and never obligatory is proposed. Because the determination of values involves several criteria, flowcharts have been constructed that display the process of selection.

Besides explaining how variables and values were defined, each section contains information about the distribution of features and values across the three samples, which helps to justify the operationalization of variables. Most importantly, this methodology chapter is a guide to understanding the annotation of the spreadsheet in the Appendix.

4.2 Language sampling

This section presents general considerations for constructing an appropriate language sample. A typology draws its best conclusions on universal trends when the languages included are as diverse as possible (Rijkhoff & Bakker 1998). The conclusions drawn based on a random set of languages might not indicate universality. Instead, a language sample is good if the languages differ genetically, geographically and structurally. Since the focus of this dissertation is to find associations between different structural properties of languages (inflection, prominence, obligatoriness), only a subset of languages qualifies for the investigation. For example, languages that do not have inflectional paradigms or contrasts in prosodic prominence cannot be included. However, it should be noted that this dissertation makes *universal* claims about the relationship between complexity and structural interaction. Thus, complexity of inflection is seen as a manifestation of general complexity, and prominent and obligatory units as a manifestation of common stabilizing structures. Justifying this universal dimension of the research question requires several methodological considerations that clarify the mediation between the general and the specific. In this regard, one has to be aware of many biases that are present when creating a language sample. In the following, I will explain the most common biases referred to in the literature and indicate solutions to minimize them. Section 4.3 will present the sample and explain to which extent these biases have been minimized.

4.2.1 Genealogical bias

Similar patterns between languages might be the result of genealogical inheritance. Genealogical bias often results from choosing languages that are prominently studied across the world, such as Indo-European languages (Bakker 2011; Rijkhoff & Bakker 1998). Thus, in order to make universal claims about language, the sample must consist of languages that are genetically so far apart that it is highly unlikely that their specific typological traits have survived unaltered. The best solution is to choose languages whose common ancestor cannot be easily determined. Within language families/stocks, it makes sense to compare relative time depth and choose those languages that share the fewest phylum nodes as possible. Rijkhoff & Bakker (1998: 269) introduce the so-called 'diversity value' that can be computed and takes into consideration "both the depth and the width of a genetic language tree." For example, Kurumanji and Scottish Gaelic are better choices for eliminating genealogical bias in the Indo-European family than English and Bavarian. The former share one node (Proto-Indo-European), whereas the latter share three (Proto-Indo-European, Proto-Germanic, Proto-West-Germanic). Linguists can access several sources available online, such as Glottolog (Hammarström et al. 2021), WALS (Dryer & Haspelmath 2013), or Wikipedia, in order to learn about the genetic relationships between the languages chosen for the study.

4.2.2 Areal proximity

The second most common bias mentioned in the literature relates to geographic proximity. Languages can share specific structural features because of contact with neighboring languages. It has been extensively shown that language contact affects the structure of neighboring languages, making them more similar (see Matras 2020 and Thomason 2001 to compare the state of the art of research). Thus, language areas ('Sprachbünde') are defined by a set of features that are present because of geographic proximity, and not because of genetic inheritance (Campbell, et al. 1986). An example of areal features is postposed definite articles among languages of the Balkans, such as Albanian, Romanian and Bulgarian (Croft 2003: 24).

This structural feature was not inherited by their common ancestor (Indo-European) but spread later across the languages of the Balkans. Another example of structural convergence is the vigesimal numerical system used in genetically unrelated Mesoamerican languages (Uto-Aztecan, Mayan, Otomanguean) (Campbell et al. 1986: 545 ff.).

Linguists that aim to reduce areal bias should choose languages from different areas of the world and be aware of features that define a linguistic area. An introduction to linguistic areas and the listing of the most relevant ones is found in Thomason (2001: 99-128). Miestamo, Bakker & Arppe (2016) suggest the creation of a "restricted sample", where every area contains the same percentage of languages from the sample. However, in certain cases, some features might be affected more by areal proximity than others, in which case knowledge about which properties define the language areas is needed. For example, choosing Basque and Spanish (both belonging to the Standard Average European area according to Haspelmath et al. 2001) in a study of morphosyntactic alignment might be still plausible, since the alignment structures are quite diverse and don't show much evidence for convergence. On the other hand, choosing Basque and Spanish for the construction of a language sample based on phonological features would inherit more bias, since the phonological inventories in both languages are extremely similar. Thus, it depends whether the typological profiles of the languages in question are influenced by contact. When doubting whether a feature is borrowed instead of inherited, one should avoid choosing languages from the same linguistic area.

4.2.3 Bibliographic bias

Biases stemming from authorship or bibliographic background were not discussed much in typological research until recently; but see Bell (1978) for an early consideration of bibliographic bias in language samples. Easterday (2017: 43–44) mentions some manifestations of bibliographic bias, such as an overrepresentation of languages that are well documented and written about (Indo-European languages, Chinese, Arabic) vs. an underrepresentation of languages where documentation is scarce or nonexistent and that have no standard form (lowland New Guinea or parts of the Amazon region) (cf. Hammarström 2010). This difference can impact the quality and reliability of sources. Written language bias can be minimized by using reference grammars compiled by trained field linguists, even from largely spoken languages. Another possibility is to minimize the number of languages in the sample that are widely spoken. Bibliographic bias is related to sociolinguistic bias since there are more reference grammars of lingua francas or standardized languages; reducing sociolinguistic bias and focusing on minority languages may also reduce bibliographic bias. Conversely, focusing on linguistic reference grammars also risks another type of bibliographic bias, namely authorship bias. Authors of reference grammars might have specific interests in a linguistic subfield. A fieldworker with an interest in morphology might give a comprehensive overview of morphological templates but not mention all syntactic constructions, and vice versa. One could minimize this bias by choosing grammars written by different authors, as proposed in Zingler (2020: 55).

4.2.4 Sociolinguistic bias

A skewed distribution of phenomena can also have sociolinguistic reasons. As noted in Chapter II, morphologically complex patterns are more likely stable in small close-knit communities with languages acquired exclusively by children, whereas languages that are spoken by a wide variety of speakers and are exchanged more often exhibit less complex patterns (Nichols 1993, Dahl 2004; Trudgill 1999, 2001, 2011; Sinnemäki 2009). Minimizing this type of bias would entail including languages with a large and wide speaker population, and including pidgin and creole languages, which are examples of high linguistic exchange. This undertaking might conflict with the bibliographical bias resulting from overrepresentation of European languages and lingua francas. Early typological studies often drew on African lingua francas such as Arabic and Swahili because those languages had reference grammars. Furthermore, choosing many Creole languages would also increase genealogical bias, since almost all Creole languages are derived from Indo-European languages.

4.2.5 Structural bias

A less-discussed bias might arise when comparing elements in languages that appear in similar structural contexts. This structural or typological bias (Bakker 2011) becomes a problem when a specific phenomenon is investigated, as is the case in this thesis. One must be aware that structural properties of languages might be the result of other, more general characteristics of these languages. For example, linguists who search common properties of suffixing languages should be sure to include languages that exhibit a different word order, since suffixing preference correlates positively with object-verb order (Hawkins and Gilligan 1988). So-called 'harmonic' relations between word order types and other structural correlations, where the order of morphemes reflects the order of separate words (Croft 2003: 62ff.; Comrie 1980; Givón 1971) should be considered before explaining other correlations. In many cases, the explanation for why one phenomenon correlates with the other is due to similar historical preconditions among languages. Minimizing structural bias can be done by consulting typological databases and contrasting the different features (Dahl 2008), such as the WALS database (Dryer &

Haspelmath 2013). This correlation between affix and word order can be demonstrated by combining the maps from Dryer (2011a) (prefix vs. suffixing inflectional morphology) and Dryer (2011b) (word order). However, Naranjo and Becker (2021) show that the structural bias between word and affix order can be minimized in a language sample large enough that is already controlled for genealogical and areal biases.

4.3 The language sample

This section describes which languages have been selected, and to which extent it was possible to minimize the different biases in constructing the sample. The overall sample consists of three smaller samples.

30 languages pertain to the main sample, that is, languages with lexically conditioned inflection (LCI). As further described in Section 4.5.1, the category of LCI comprises lexically conditioned verbal inflectional allomorphs, syncretic morphs, and segment sequences involved in stem alternation. The languages of the LCI sample can also have grammatically (GCI), phonologically (PCI) and unconditioned inflection (UCI). Grammatically conditioned inflection includes inflectional allomorphs or syncretic morphs that are conditioned by other grammatical categories, as well as cumulative morphs (see Section 4.5.2). Phonologically conditioned inflection includes allomorphs that are conditioned by the phonological environment alone (morph variants). These have been grouped together with unconditioned inflection (UCI) which includes inflection that does not have allomorphy or is not syncretic or cumulative, except for indexation morphs that express person, number, predictable gender and clusivity cumulatively, but refer to only one referent (e.g., subject) (see Section 4.5.3).

The remaining languages belong to control samples: 6 languages which do not have LCI but do not have grammatically conditioned inflection (GCI) and 5 languages which have neither LCI nor GCI but either phonologically conditioned inflection (PCI) or unconditioned inflection (UCI). As indicated in Chapter II and further justified in Section 4.4.1, inflectional morphology was chosen to study the distribution of difficulty-based complexity resulting from morph conditioning because inflectional categories have a similar degree of semantic generality, a control property that derivational or lexical morphology does not exhibit. The distinction between inflectional and non-inflectional categories is determined in Section 4.4.1.2.

Figure 4.1 shows a world map with the 41 languages chosen for the study. The dots are located relative to the center of where these languages are spoken, as determined by WALS (Dryer and Haspelmath 2013) and Glottolog (Hammarström et al. 2021).

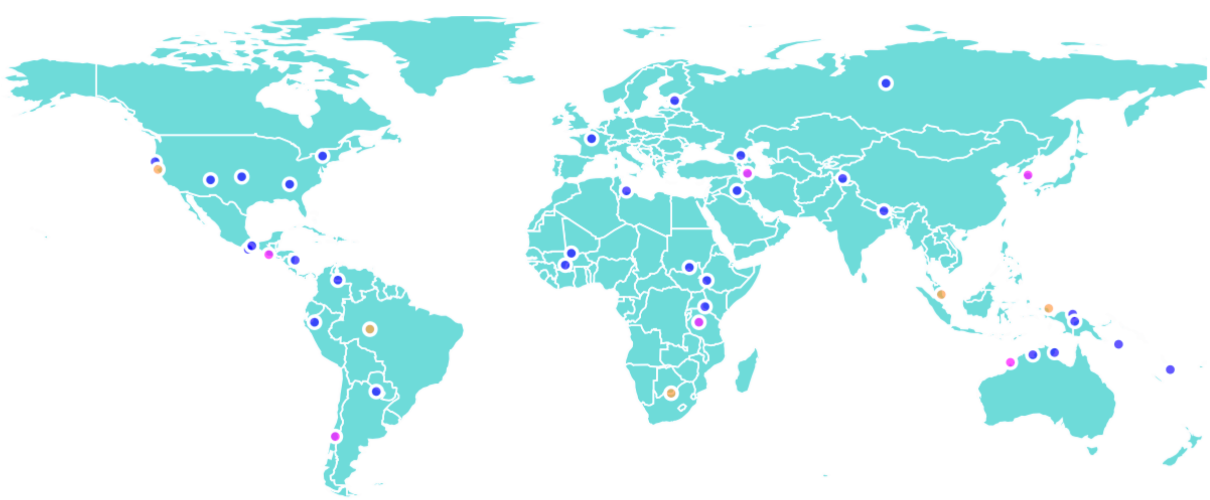


Figure 4.1: Location of the languages chosen for the three samples (Blue: LCI sample; Magenta: GCI sample; Orange: UCI sample).

The following subsections explain how each sample was constructed and how the selection of languages minimize sampling biases as presented in Section 4.2.

4.3.1 LCI sample

The main sample was filtered for the property 'lexically conditioned inflection.' This reduced vastly the languages that could be chosen. Since there is no database for languages that exhibit LCI, much time has been spent on researching individual grammars for this property. Investigations on lexically conditioned conjugation classes are also rare. Some non-Indo-European examples include Blevins (2007) for Estonian or Blevins (2005) for Yurok. The GRAMCATS database (hosted at the University of New Mexico, described in Bybee, Perkins and Pagliuca 1994) exhibits morphs annotated for type of conditioning (including lexical conditioning), which helped in the first stage. However, the GRAMCATS database lists selected morphs and not all the verbal paradigms of the language. Making the sample as diverse as possible was the priority in order to minimize genetic bias. This prevented the inclusion of cognate morphs; since LCI morphs are usually older than non-LCI morphs, it is likely that languages will have cognate morphs among the LCI morphs.[10] Some language families and branches with many languages and speakers seem to lack LCI (Bantu, most of the Austronesian branches, Trans-Eurasian[11], Dravidian, Austroasiatic, Sinitic). This does not mean that there are no languages

---

[10] This has not been investigated; however, I assume that LCI morphs are older because they tend to be closer to the lexical stem, which reflects an older age (Mithun 2011). In addition, one could argue that lexical conditioning is a manifestation of mature morphology which correlates with age (Dahl 2004).

[11] Trans-Eurasian is formerly known as Altaic; According to Robbeets and Savelyev (2020), the term Trans-Eurasian includes Turkic, Mongolic, Tungusic, Japonic, and Koreanic, whereas Altaic does not always include Japonic and Koreanic. The term also serves to eliminate the reference to the Altai mountains as a potential homeland. I do not intend to argue

with LCI in these families. In order to reduce the length of time spent in searching for LCI languages, the search was stopped if four languages from different branches of one family did not exhibit LCI. In the end, only one language was chosen from each family, with some exceptions that are explained in the following sections.

The priority to minimize genetic bias resulted in some areal shortcomings. For example, the LCI sample lacks languages from Southeast Asia. From the Austronesian language family, only some languages in Vanuatu exhibit LCI. Inflection is generally lacking in the languages of Southeast Asia. The search for LCI in East and Central Asia was also unsuccessful. The reason is that most of these language families (Japonic, Korean, Mongolian, Turkic, Tungusic, Nivkh, Chukchi, Yukagir) have regular, agglutinating inflection and do not organize inflection in lexical classes nor exhibit stems alternating for inflectional categories. The same holds for languages from the Amazon, Central and Southern Africa (Bantu and Khoisan). The American Northwest and Canada are also areas with a scarcity of LCI, except for the Athabascan family.

The grammars of the selected languages were all written by different authors, apart from Nunggubuyu and Bangime, where Jeffrey Heath was author and co-author, respectively. Several languages in the sample belong to small families or are isolates (such as Ket, Betoi, Burushaski, Sumerian, Yelî Dnye, Bangime, Pirahã). Isolates are over-represented in the sample, since many of these show LCI, while surrounding families do not (e.g., Burushaski, Betoi, Sumerian, Yelî Dnye). However, since the languages are maximally genetically distant from one another, the selection of several isolates was necessary. From a phylogenetic perspective,

whether Altaic or Trans-Eurasian is a better name for the family, nor do I claim that Trans-Eurasian is a valid hypothesis; this concept is used for typological purposes.

the sample does not excessively represent isolates, as isolates account for more than a third of the world's primary language phyla (162 out of 406, as estimated by Campbell 2018, with a conservative approach that avoids merging these phyla into larger groups).

Despite exhibiting LCI morphs, the languages are morphologically and syntactically diverse. There is an approximately equal distribution of prefixes/proclitics (52 %) and suffixes/enclitics (48 %) in the sample. Given the global suffixing preference (Sapir 1921: 70; Dryer 2011a), one can say that prefixing languages are more likely to exhibit LCI. With regards to word order, LCI languages show a predominance for verb-finality. The sample counts 17 SOV languages (54.8 %), which is more than the global proportion – 41.0 % according to WALS (Dryer 2011a). Verb-medial languages are relatively underrepresented in the sample, namely 22.6 %, in contrast to the global proportion of 35.5 % (Dryer 2011a). As Yadav et al. (2020) show, languages with a predictable subject-object-verb order are more likely to tolerate morphological richness, of which LCI is certainly a symptom. This is related to another peculiar observation, namely that there are no LCI languages that have a single inflectional position. LCI is a mature phenomenon (Dahl 2004) that tends to coexist with more predictable morphology, and languages without bound morphology do not show lexical conditioning of grammatical morphs. As such, the sample exhibits languages with at least two inflectional verbal positions (this is the case for Estonian). Nevertheless, this peculiarity represents an advantage for examining the intra-language correlation of conditioned morph types and the structural variables and reduces the need to include even more languages into the sample. On average, LCI languages had 6.7 inflectional positions (including clitics),[12] reflecting their high

[12] As stated in Section 4.4.2.2, a position is considered inflectional if it exhibits at least one morph that is inflectional. Thus, the proportion of the numbers of inflectional and non-

degree of synthesis of the verb (cf. Bickel & Nichols 2013). The 30 languages of the LCI sample provided 316 positions to analyze (341 if including sub-positions which are explained in Section 4.4.2.3), which is enough to achieve significant results in the correlation between inflectional morphs and the structural variables prominence and obligatoriness.

In conclusion, the LCI sample generally fulfills the criteria of being a balanced sample with minimal genetic, areal, bibliographic and structural bias. In the following, the languages are presented grouped by their macro-area. The definition of macro-areas differs among authors; however, the most common typological sources (WALS and Glottolog, which however are not independent sources) count six areas (Hammarström et al. 2021): Africa, Eurasia, North America, South America, Australia and Papunesia. Similar to Zingler (2020), I subsume Australia and Papunesia into one linguistic area, 'Oceania,' for reasons which are mentioned further below. The following sections will also discuss the challenges in extracting the information needed to annotate the languages in the spreadsheet.

4.3.1.1 Africa

Africa is the macro-area that contains the largest number of the world's languages. According to the high estimate of Glottolog (Hammarström et al. 2021), 2353 (about one third of the world's languages) are spoken in Africa. However, this macro-area also contains the fewest linguistic families, which represents a challenge for genetic stratification. Depending on classification, the number of families ranges from 6 to 15. In addition, not all families contain languages with the same degree of morphological complexity. Lexically conditioned inflection

---

inflectional positions is greater than the proportion of the numbers of inflectional and non-inflectional morphs.

is abundant in Afroasiatic languages but absent in Bantu and the so-called ‘Khoisan’ families. Morphological complexity is particularly present in Eastern Africa (Sudan, Kenya, Tanzania, Ethiopia), across language families (Afroasiatic, Niger-Kongo, Nilo-Saharan). Therefore, three of the six LCI languages (Sheko, Lumun, Dholuo) have been chosen from this area. Four other languages are in the same family as another language in the sample (Lumun & Supyire, Zuaran Berber & Sheko), but they are in Western and Northern Africa. Due to the small number of language families, Africa is the only macro-area where the genealogical bias based on the concept of a family could not be fully minimized. However, it has been ensured that these related languages are still genetically and areally distant from one another. According to Rijkhoff and Bakker (1998), what matters is the diversity of the language family, i.e., dispersion among the branches; and the ‘diversity value’ for African language families is higher than the one of families in Eurasia, for example. Supyire and Lumun pertain to the Niger-Congo family, but they belong to different sub-branches (Senufo and Talodi-Heiban, respectively, which are two primary branches), and are spoken far from one another (Western and Eastern Africa, respectively). Sheko and Zuwara Berber are from the Afroasiatic family, namely Omotic and Berber, again two primary branches. Geographically, Zuwara Berber is spoken in Tunisia and Sheko in Ethiopia. Furthermore, one language isolate (Bangime) has been included. All grammars contained the necessary information for annotating the categories under investigation. For Bangime, two different grammars were consulted (Hantgan 2013; Heath & Hantgan) to obtain a better picture of the paradigms and their obligatoriness. The grammar of Zuwara Berber (Mitchell 2009) did not include glosses but provided precise information on the conditioning of the morphs, as well as the stress patterns. The languages chosen for Africa are listed in Table 4.1.

Language	***ISO 639-3***	***Family***	***Genus***	***Location***	***Bibliographic source***
Bangime	dba	Bangime	Bangime	Mali (Dogon Cliffs)	Hantgan (2013); Heath & Hantgan (2018)
Lumun	lmd	Niger-Kongo	Talodi-Heiban	Sudan	Smits (2017)
Luo	luo	Eastern Sudanic	Nilotic	Lake Victoria	Tucker and Creider (1994)
Sheko	she	Afro-Asiatic	Omotic	Ethiopia	Hellenthal (2010)
Supyire	spp	Niger-Kongo	Senufo	Mali/Ivory Coast	Carlson (1994)
Tunisian-Zuwara Berber	tuni1262 (glottocode)	Afro-Asiatic	Berber	Tunisia	Mitchell (2009)

Table 4.1: LCI Languages from Africa.

4.3.1.2 Eurasia

Eurasia is, in geographical terms, the largest macro-area, and contains 1977 languages according to Glottolog (Hammarström et al. 2021). Eurasia offers more language families to choose from than Africa, although the diversity value of these families is lower than the one of families in Africa (Rijkhoff and Bakker 1998: 272). However, not all language families have inflection, and some of the languages have very little morphology. This is the case in East and Southeast Asia. Languages from Central Asia and Southern India do not seem to exhibit LCI either. These languages have an agglutinative morphology, and agglutination is less likely to be morphologically (grammatically or lexically) conditioned (Zingler 2020: 21-22). All languages chosen from Eurasia are maximally genetically distinct. Two isolates were included, Burushaski and Sumerian, the latter of which is recorded from ancient cuneiform scripts. With regards to bibliographic quality, all grammars and descriptions provided the necessary information on inflectional conditioning contexts, prosodic prominence and morphological obligatoriness. For the construction of a template of Standard French, Ashby's (1977) analysis has been used, since this analysis is more templatic than the conventional descriptions found in French. For determining the nature of LCI in Estonian, the grammatical description of conjugation classes by Blevins (2007) was consulted. The languages chosen from Eurasia are presented in Table 4.2.

Language	*ISO 639-3*	*Family*	*Genus*	*Location*	*Bibliographic source*
Burushaski	bsk	Burushaski	Burushaski	Pakistan/India	Berger (1998)
Dumi	dus	Mahakiranti	Sino-Tibetan	Nepal	Van Driem (1993)
Estonian	est	Uralic	Finnic	Estonia	Blevins (2007), Harms (1962)
French (Standard)	fra	Indo-Euro-pean	Romance	France	Batchelor & Chebli-Saadi (2011), Ashby (1977)
Ingush	inh	Nakh	Nakh-Dage-stanian	Russia (North Caucasus)	Nichols (2011b)
Ket	ket	Yeniseian	Yeniseian	Russia (Sibe-ria)	Vajda (2004), Werner (2002)
Sumerian	sux	Afro-Asiatic	Berber	Mesopotamia	Jagersma (2010)

Table 4.2: LCI languages from Eurasia.

4.3.1.3 North-America

The sample of North America is geographically unbalanced in two ways: it contains the same number of languages than Eurasia despite being a smaller area, and second, no language from the Northern part of North America has been included in the sample. Besides the Athabascan language family, there seems to be a lack of LCI in Eskimo-Aleutian, Algonquin, Salishan, Wakashan and some Californian languages. The reader might wonder why a different Athabascan language was not chosen that could fill the empty area in Alaska or Canada. There are three reasons for choosing Navajo from the Southwest instead. First, there is evidence that the Yeniseian languages might be related to the Na-Dene languages in North America (most prominently argued by Vajda 2010). By choosing one of the geographically most distant relatives in the Southwest, this reduces the geographic bias between Dene and Yeniseian. Nevertheless, the genealogical relationship between these languages is very distant if there is a relationship at all. Flegontov et al. (2016) dates the split of Na-Dene and Yeniseian around 6,500–7000 years before present. Second, it is hard to find information on either the conditioning of each morph across Athabascan languages and stem alternations, or on the existence of stress. Rice's (2000) grammar on Slavey, despite exhibiting detailed information on the conditioning of inflection, could not be used since stress is only described for the Hare dialect (Rice 2000: 46), making it difficult to generalize across the morphological paradigms that are not always from Hare. Third, my own expertise in Navajo made it easier to research the correlation between the morphological complexity, prominence and obligatoriness.

Towards the Southern part of the continent, two languages from Mexico have been selected that exhibit LCI: Ayutla Mixe and Eastern Chatino. The areal proximity within Mesoamerica could not be reduced due to the lack of LCI in other language families (such as Mayan and Uto-Aztecan) and the lack of comprehensive description of other languages. For Eastern Chatino, a description of the morphological and phonological system was sufficient to annotate the different positions of the template and the alignment with prominence and obligatoriness. Grammars of other Zapotec languages lacked a description for one or more properties. The selected languages for North America are listed in Table 4.3.

Language	*ISO 639-3*	*Family*	*Genus*	*Location*	*Bibliographic source*
Ayutla Mixe	mxp	Mixe-Zoque	Mixe-Zoque	Mexico (Oaxaca)	Robero-Mendez (2009)
Kiowa	kio	Kiowa-Tanoan	Kiowa-Tanoan	USA (Oklahoma)	Watkins (1980, 1984)
Koasati	cku	Muskogean	Muskogean	USA (Lousiana, Texas)	Kimball (1991)
Navajo	nav	Athabascan	Na-Dene	USA (Arizona, New Mexico)	Young & Morgan & Midgette (1992); McDonough (2003)
Oneida	one	Northern Iro-quoian	Iroquoian	USA (New York, Wisconsin) Canada (Ontario)	Abbott (2000)
Yurok	yur	Yurok	Algic	USA (Northwest-ern California)	Robins (1958); Blevins (2003; 2005)
Zacatepec Chatino	ctz	Zapotecan	Oto-Manguean	Mexico (Oaxaca)	Villard (2015)

Table 4.3: LCI languages of North America.

4.3.1.4 South America

South America is the least represented area in the LCI sample. Despite being the genetically most diverse continent (exhibiting over 100 families/isolates), I have found few languages/language families that exhibit LCI. In particular, I have not found a grammar from a language spoken in the Amazonian basin that exhibits LCI. Most of South American languages have an agglutinating verbal morphology, without lexical conjugation classes. In addition, many languages from South America lack a comprehensive description. Following Dryer's definition (1989: 268), which sets the border between North and South America around Honduras (1989: 268), I included the Central American language Ulwa (Misumalpan, Nicaragua) into the South American sample. In total, the LCI sample exhibits 4 languages from South America. The Southern Cone and the Amazon were chosen for languages of the control sample (Mapudungun and Pirahã respectively). All descriptions of the languages were sufficient to inform the coding of the verbal positions. However, the "paucity of data" in Betoi, an extinct language, is also reflected in the author being unable to determine the conditioning of some morphs (cf. Zamponi 2003: 29). The languages chosen from South America are listed in Table 4.4.

Language	***ISO 639-3***	***Family***	***Genus***	***Location***	***Bibliographic source***
Aguaruna	agr	Jivaroan	Jivaroan	Perú	Overall (2017)
Betoi	beto1236 (glottocode)	Betoi-Jirara	Betoi-Jirara	Venezuela	Zamponi (2003)
Pilagá	cku	Guaicuruan	South Guai-curuan	Argentina	Vidal (2001); Klein (1973)
Ulwa	ulw	Misumalpan	Misumalpan	Nicaragua, Honduras	Green (1999)

Table 4.4: LCI languages from South America.

4.3.1.5 Oceania

For this investigation, it has been decided to group together the macro-areas Australia and Papunesia (as distinguished in Glottolog) under the macro-area 'Oceania.' The reason is that the features of Australia's language families are quite similar, and this continent might represent a linguistic area of its own (Dixon 2001). Most of the genetic diversity exists in Northern Australia, whereas the Pama-Nyungan family spreads over the rest of the continent. Much of the linguistic similarity resides in the structure of verbs. For example, Nunggubuyu and Murrinh-Patha belong to distinct language families but share similarities in the grammatical categories and the positions in the verbal template, such as a prefixal position with cumulative tense/mood/aspect/indexation morphs. It is stated that the intense and millennia-long lasting contact between distinct Australian language families had a deep impact on evening out phonological, semantic and morphosyntactic differences (Dixon 1997, 2001; Miceli and Dench

2017). As such, only two languages from Australia have been selected, and none from the Pama-Nyungan family, since LCI seems to not be present there.

Languages from Papunesia, on the other hand, exhibit less evidence of areal contact. Due to the high genealogical and typological diversity, four languages from Papunesia have been included in the sample (Skou, Mian, Yele, Ura). Ura, spoken in Vanuatu, is one of the few Austronesian languages that has LCI. Austronesian languages are usually either rich in derivation (Taiwan, Philippines) or have no bound morphology at all (Polynesia). The grammatical descriptions of the languages from Oceania are comprehensive enough for annotating them in the spreadsheet. These languages are presented in Table 4.5.

Language	***ISO 639-3***	***Family***	***Genus***	***Location***	***Bibliographic source***
Mian	mpt	Trans-New-Guinea	Ok	Papua New Guinea, Sand-aun Province	Fedden (2011)
Murrinh-Patha	mwf	Southern Daly	Murrinh-Patha	Australia, Northern Ter-ritory, Wadeye	Walsh 1976; Nordlinger & Caudal (2012); Mansfield (2017)
Nunggubuyu	nuy	Gunwinyguan	Nunggubuyu	Australia, Northern Ter-ritory, Num-bulwar	Heath (1984)
Skou	skv	Skou	Western Skou	Indonesia, Muara Tami District	Donohue (2004)
Ura	uur	Austronesian	Oceanic	Vanuatu, Er-romango	Crowley (1999)
Yelî Dnye	yle	Yele	Yele	Papua New Guinea, Ros-sel Island	Henderson (1995)

Table 4.5: LCI languages of Oceania.

In the following, the control samples are presented, and their construction explained.

4.3.2 Control samples

Since the examination in this dissertation targets a specific trait/construction (i.e., LCI) in certain contexts (verbs) without analyzing the context in which this traits occurs, the typology is 'non-holistic' (Himmelmann 2000). A holistic typology would instead involve an in-depth study of lexically conditioned inflection, across different types of constructions. Furthermore, LCI is not studied in isolation, but rather in combination with GCI and UCI with regards to morphological complexity. However, since LCI, GCI and UCI cover all types of inflection, this is a typology of inflection in general.

While LCI languages provide enough data for GCI and UCI, one cannot arrive at universal conclusions for the behavior of GCI and UCI by only looking at how they behave in LCI languages. Indeed, the presence of LCI might have effects on the distribution of GCI and UCI. For example, if in one language, stress falls consistently on LCI positions, a difference in stress placement between GCI and UCI cannot be determined. Therefore, the correlation between GCI and UCI must be investigated in a sample of languages that do not exhibit LCI. Conversely, GCI languages might not provide an ideal insight for the correlation of UCI and non-inflectional positions. That is, in order to control for the structural impact that GCI has on UCI and non-inflectional positions, an additional sample of languages that has neither GCI nor LCI is needed. I will call this the UCI sample, even if it contains phonologically conditioned inflection. In the end, a GCI sample consisting of 6 languages and a UCI sample of 5 languages have been included to control for the effect of having different types of conditioned inflection in one language.

4.3.2.1 GCI sample

The GCI sample consists of languages whose verbal constructions exhibit grammatically conditioned inflection of the type discussed in Chapter II (Section 2.4.2). One language from each macro-area was selected, except for Eurasia. For Africa, I chose a Bantu language, Kinyamwezi, which has a grammar whose author provids a verbal template and sufficient information on morph conditioning. For Eurasia, I decided to select agglutinating languages since this type of synthesis was underrepresented in the LCI sample. However, finding appropriate GCI languages from this area was more difficult than I expected. First, it was not easy to find a good English description of a Central Asiatic language. Another challenge was that authors did not necessarily provide a template of morpheme order. The comprehensive grammar of Korean (Sohn 1999) filled the geographical gap in Eastern Asia, and Iranian Azerbaijani was chosen for Western Asia.

With regards to North America, I choose a language with not as many verbal positions so that structural bias resulting from polysynthesis could be minimized. The Mayan language K'ichee' was an appropriate candidate whose verbal constructions contain GCI and whose description (Mondloch 2013) contains enough information to deduce the alignment of the morphological positions with prosodic prominence and obligatoriness. Conversely, I chose Mapudungun as a language with many verbal positions from South America. From Oceania, I selected Bardi, a language of the Nyulnyulan family, the Westernmost Non-Pama-Nyungan family. The languages of the GCI sample are presented in Table 4.6. Most of the inflectional positions of this sample are suffixing (34 suffixing; 11 prefixing), mostly due to the input from Mapudungun, which contributed 15 inflectional suffix positions.

Language	***ISO 639-3***	***Macro-area***	***Family***	***Genus***	***Location***	***Bibliographic source***
Azari (Iranian)	azb	Eurasia	Altaic	Turkic	Northwest Iran	Lee (1996)
Bardi	bcj	Oceania	Nyulnyulan	Nyulnyulan	Australia, Western Australia, Dampier Peninsula	Bowern (2004)
K'ichee'	quc	North America	Mayan	Mayan	Guatemala	Mondloch (2013)
Korean	kor	Eurasia	Korean	Korean	Korea	Sohn (1999)
Mapudungun	arn	South America	Araucanian	Araucanian	Southern Chile	Smeets (2008)
Nyamwezi	nym	Africa	Niger-Kongo	Bantoid	Tanzania	Maganga & Schadeberg (1992)

Table 4.6: GCI language sample.

4.3.2.2 UCI sample

The second control sample consists of languages that exhibit neither LCI nor GCI but do have UCI. Because there is not enough evidence that suggests a difference in complexity between phonologically conditioned and unconditioned inflection (Chapter II), I have grouped these two types into this sample. Indeed, it is hardly possible to find enough languages whose morphology is not phonologically conditioned (i.e., no morphophonology). Furthermore, some areas of the world appear to lack languages where inflection is merely phonologically conditioned. One of the few such languages in North America is Northern Pomo, with transparent tense/mood/aspect suffixes. Likewise, Pirahã from the Amazon exhibits inflectional morphs that are not conditioned by grammatical or lexical elements. Because Pirahã and Northern Pomo contain only inflectional suffixes, other languages (Hatam, ǂHȍã and Semelai) were chosen that have inflectional prefixes, balancing the sample for both types of morph positions (11 prefixing and 14 suffixing inflectional positions in total). Semelai is the only Southeast Asian language in the overall sample, and one of the few Austroasiatic languages that has inflectional morphs. Hatam, a language from the Bird's Head in New Guinea, is different from other Papuan languages by barely having affixation. The UCI sample is presented in Table 4.7.

Lan-guage	***ISO 639-3***	***Macro-area***	***Family***	***Genus***	***Location***	***Bibliographic source***
ǂHòã	huc	Africa	Kxa	ǂHoan	Botswana	Collins & Gruber (2014)
Semelai	sza	Eurasia	Austroasiatic	Aislian	Malaysia, Malay Pen-insula	Kruspe (2004)
Northern Pomo	pej	North America	Hokan	Pomoan	United States, Northern California	O'Connor (1987)
Pirahã	myp	South America	Mura	Mura	Brazil, Ma-ici River	Everett (1986)
Hatam	had	Oceania	West Papuan	Hatam	Papua New Guinea, Eastern Bird's Head	Reesink (1999)

Table 4.7: UCI language sample.

4.4 Morphological annotation

Chapter II examined which parameters and variables that show degrees of morphological complexity can be constructed for typological examinations. Based on psycholinguistic and information-theoretical evidence, it was suggested that the tripartite distinction of the dependent variable of inflectional complexity into unconditioned/phonologically conditioned, grammatically conditioned and lexically conditioned morphs not only reflects a degree of complexity but gives typological linguists a concrete guide in the search through grammatical descriptions. On the other hand, it was suggested that the independent variables prosodic prominence and morphological obligatoriness are promising structural factors that can be investigated as they are associated with structural retention. In order to annotate these variables, an outline is needed that captures the morphological and prosodic dimension of these categories. The main morphological distinction is between inflectional and non-inflectional morphs. Section 4.4.1 describes which morphs have been selected as inflectional, to be further annotated for inflectional conditioning, and which morphs have been excluded. Section 4.4.2, on the other hand, describes how positions have been established to allow the annotation for prosodic prominence and obligatoriness.

4.4.1 Determining inflectional categories

Since this dissertation makes assumptions about the processing of inflectional morphs, it is important that these morphs are, with exception of their degree of complexity, controlled for other features that influence processing. Two factors that might influence ease of processing are semantic content and frequency of occurrence. These two factors are mostly apparent in the studies on the trichotomy between inflection, derivational and stem morphs (see Leminen et al. 2019 for a review on these effects). According to Bybee (1985), inflectional morphs tend to express more general semantics, and derivational morphs more specific semantics. On the other hand, inflectional morphs are less relevant, and derivational morphs more relevant for the meaning of the stem. With regards to frequency, inflectional morphs tend to occur more frequently both in type and token frequency. Derivational morphs are more restricted in their use and have a lower type and token frequency.

Furthermore, inflectional morphs are more consistent in their semantic properties and frequency of occurrence. Tense, mood and some aspect categories have a similar degree of generality and are frequently repeated in the discourse. The semantics of derivational morphs are more varied and dependent on the specific semantics of the lexemes, which can be quite diverse. For example, lexemes can be specific like ‘chopstick’ or general like ‘environment.’ Derivational categories might entail specific modifying categories, such as ‘in the same way’ or ‘mediopassive’ (the latter which can be applied to verbs with certain semantics), or be quite general, such as ‘transitive.’ However, transitivity is rarely overtly expressed in verbs, and therefore, most derivational morphs have less general meanings. Inflectional categories, on the other hand, are mostly general, such as ‘future,’ ‘subjunctive’ or ‘third person.’ Because inflectional categories are often obligatory in verbal constructions, the frequency of inflectional categories is expected to be more similar than of derivational categories.

The semantic and discursive similarity of inflectional categories lends itself to the comparison of more specific differences. As such, inflectional morphs are a good choice to investigate the different types of conditioning. Verbs usually comprise most of a language’s inflectional morphology, in terms of degree of synthesis and number of grammatical categories. For example, adjectives usually index referents and exhibit morphs denoting degree (such as comparative and superlative); nouns can exhibit case, number, definiteness and gender; verbs, in turn, comprise even more categories like valency, tense, mood, aspect, polarity, indexation (subject, object, indirect object) or evidentiality. Verbs therefore provide more options for the investigation of complex inflection. However, ‘inflectional’ is not always a concrete feature that can be looked up in grammars and separated from derivational features. While every grammatical description references a verb or 'verbal complex', the morphological categories that a verb has can vary quite a lot. The distinction between derivation and inflection can be understood as a gradual cline (Bybee 1985; Dressler 1989; Plank 1994), although other linguists disagree (Perlmutter 1988; Anderson 1982; 1992; see Booij 2006: 659). Nevertheless, one can generally trust grammars that those categories deemed inflectional are the most general and frequent morphological categories of a language.

4.4.1.1 General criteria

What is important for this study is the semantic and frequency-based aspect of inflection. This means, grammatical categories should be selected that

a. are frequent
b. exhibit a similar type frequency
c. are semantically general (closely interconnected with b.)

But how can typologists ensure that they chose the categories based on these criteria, when the grammar does not indicate which categories are general and frequent in a language? There may be some solutions to this problem.

According to Greenberg (1966), discourse frequency of conceptual inflectional categories is similar across languages. This is reflected in high semantic generality of a category, and therefore, in high typological frequency, that is, how often a specific category is found in the languages of the world. In order to select general morphs, one could rely on the morphological categories that are frequent in grammatical descriptions, such as tense distinctions (present/past/future), person (first/second/third), or number (singular/plural; although number is also expressed through less inflectional morphology such as pluractional morphs). If a language makes a distinction according to a cross-linguistically common category, the morphs expressing the features of this category are likely to be general and frequent in discourse. Of course, some languages exhibit frequent and general categories that are typologically rare. One example are honorifics in Korean which are very often part of utterances. These honorifics are a typological peculiarity of South and East Asia. On the other hand, grammatical gender is found predominantly in languages of Europe and Africa but is rare in East Asia or the Americas.

Second, detecting high generality and frequency can be induced from the existence of paradigms. If a category exhibits different morphs with distinct features, the category that comprises these morphs will be present in every word that exhibits one of these morphs. If the category is expressed by a free adverbial word, it is likely that there are fewer constructions based on this adverbial. Thus, paradigmatic tables are a hint for determining whether one category is general or not in the language. The property of a category exhibiting multiple features is also referred to as behavioral potential (Croft 2003), which is correlated to frequency, in the way that the unmarked member within the paradigm has a higher text frequency. For example, if a category has different functional elaborations, such as recent, distant and hesternal past, then it is plausible to assume that (at least some) past morphs have a high frequency in the language, and as such, the category 'tense' is highly entrenched in the minds of speakers.

In conclusion, the selection of general and frequent grammatical categories (what should be called inflectional in this study) can be assessed by targeting morphs whose features are typologically frequent and/or are organized in paradigms of frequent categories. The specific criteria for distinguishing inflectional from non-inflectional categories is explained in the following.

4.4.1.2 Specific criteria

Based on the properties for highly general and frequent morphological categories presented in Section 4.4.1.1, several features from the categories tense, mood, aspect, indexation and polarity were included to investigate the relationship of inflectional complexity and structural properties. I excluded common derivational categories such as valency operations (reflexive, transitivizer, applicatives, passive/antipassive markers) because of their likelihood to be too relevant to the lexical meaning of the verb – that is, being less semantically general, or exhibiting non-consistent degrees of generality.

With regards to inflectional categories, I proceeded by integrating both typological commonality and paradigmatic properties in the selection process. The categories gathered from the languages of the sample were sometimes frequent, sometimes infrequent. The inflectional categories that are most commonly encountered in cross-linguistic research are listed in (1).

(1) Most common/general inflectional (TAMIP) features

Tense: present/past/future

Aspect: imperfective, perfective (or: telic/atelic; bounded/unbounded). According to Bybee (1985), this distinction is the most general aspectual semantic contrast, but there are other common ones.

Mood and Modality: epistemic (realis/indicative, irrealis/subjunctive); deontic (imperative, potential, optative/desiderative); interrogative (polar question, content question)

Indexation: person (1, 2, 3), number (sg, du, pl), gender/class (animate, inanimate; feminine, masculine)

Polarity: negative

Whenever I found that a language exhibited a common inflectional feature, I annotated it as inflectional. If a language did not exhibit the features listed in (1), several further steps related to paradigmatic relations followed in the selection process. This process is illustrated in the flowchart in Figure 4.2.

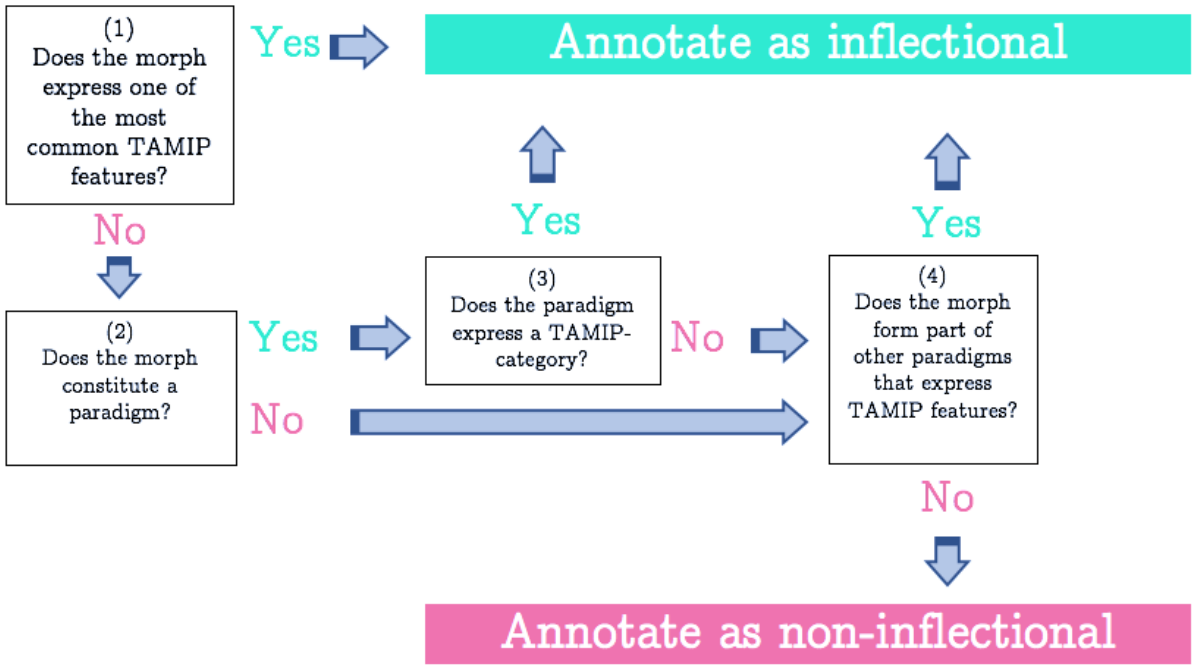


Figure 4.2: Flowchart for choosing whether a morph should be annotated as inflectional or not. Both high generality and behavioral potential (paradigmaticity) are considered to determine appropriate morphs. The elimination process can be exemplified in five languages.

K'ichee' *in-* '1sg.absolutive' inclusion after phase 1

- *in-* expresses common TAMIP-features (first person, singular).
- *in-* is annotated as inflectional.

Korean pluperfect 'past-past' *assess*: inclusion after phase 3

- Pluperfect is not a common TAMIP-feature, instead, the semantics are peculiar.
- However, *assess* is part of the paradigm with *ass* (past/present perfect) (position 3, see spreadsheet).

- *–assess* and *–ass* pertain to a TAMIP-category (tense).
- *assess* is annotated as inflectional.

Semelai *ga=* 'imminent aspect': inclusion after phase 4

- 'imminent' is not a common TAMIP-feature.
- *ga=* does not form a part of a paradigm in the language; it occupies a single position, namely in P-4 (see spreadsheet).
- *ga=* acquires future meaning in imperfective constructions. Both future and imperfective are common TAMIP-features.
- *ga=* is annotated as inflectional.

Navajo: *n-* 'spherical object': exclusion after phase 4 (1 no – 2 no – 4 no)

- 'spherical object' sounds like gender/class but is not found among the most common TAMIP features.
- *n-* appears in the same position as other morphs like and *n-* 'mind' *n-* 'eye' and *d-* 'fire' or *d-* 'speech', but this can't be called a paradigm since there is no overarching grammatical category (e.g., gender/class) for the meanings expressed by these morphs.
- *n-* does not form part of other TAMIP-paradigms, it is rather a lexical requirement of some verbs (Young, Morgan & Midgette 1992: 853)
- *n-* is annotated as non-inflectional.

Koasati: *-v́:mo* 'intensely': exclusion after phase 4 (1 no – 2 yes – 3 no – 4 no)

- 'intensely' is not a common TAMIP-feature.

- *-v́:mo* appears in a paradigm with *-á:ho(:)s(i)* (very) *-báhno* (must; be obliged) *-fihn(a)*/*-fį*/*-fin* (too much) *-hónk(a)* (really) *-má:li* (in the same way) *-mbí:k(a)* (a pleasant deal) *-v́:mo* (intensely) *-ná:n(a)* (all the time) *-palámmi* (extremely).
- The paradigm does not express a TAMIP-category, the meaning of the morphs has semantics related to degree of a property or action.
- *-v́:mo* does not co-occur in other paradigms
- *-v́:mo* is annotated as non-inflectional.

By using this chart, several categories and features have been included in the investigation. The following sections contain a detailed summary of these.

4.4.1.3 Included categories and features

Several features were annotated in the 41 languages as belonging to the categories tense, mood, aspect, indexation or polarity. The following subsections inform about the name and count of features across the languages. It has been attempted to categorize the general categories (e.g., tense) into 'categories' (e.g., future/past) and features. However, both the categories as well as the features might be different between languages, and not every language exhibits even the most general categories.

4.4.1.3.1 Tense

The selected tense categories are presented in Table 4.8.

Categories	Features	Count (in languages)
Past	past, immediate p., recent p., intermediate/distant p., remote p, narrative p., hesternal p., non-hodiernal p., near p., sequential p., relative p., normal p., prior p., dependent p., pluperfect p., past-past, I p., II p., III p., IV p.	31
Future	future, immediate f., definite f., posterior f., distant f.	22
Other	aorist	5
Non-Past	non-past, non-realized n-pa., I, II, III n-pa.	3
Non-present	proximate n-pr., remote n-pr.	1

Table 4.8: Tense categories and features across languages. Present not included since it is not usually discussed in grammars and is often zero-marked. The name of categories function as features in languages (present, past, future), when there is no further specification. The features can be co-expressed with other non-tense features; however, these other features are not mentioned here.

Table 4.8 shows that languages of the sample have more features related to past tense, whereas distinctions like non-past or non-present are rare. The count shows how many languages exhibit a specific main type. Present is not annotated due to its unmarked status in many languages and is as such not often elaborated in the grammars (however, it is expected that every language has a category for present if it has a category for past or future). Only few languages (such as Burushaski and Ket) have overt present markers, but these languages also exhibit other marked tense distinctions. The feature 'Aorist' does not belong to a uniform category, but it is included under tense, although it co-expresses aspect. For example, Aorist in Azari expresses (indefinite) future events (further function: habitual) (Lee 1996: 48), whereas Aorist in Oneida marks past time (further functions: factual, progressive, translocative) (Abbott 2000: 15). All the specific subtypes have been annotated as inflectional because they fulfilled the criteria in phases 2, 3 and/or 4 (Figure 4.2).

4.4.1.3.2 Mood and modality

The annotated mood and modality categories/features are presented in Table 4.9.

Categories	Features	Count (in languages)
Epistemic	ability, affirmative, apprehensive, attemptative, certainty, contrafactual, declarative, dubitative, factual, impossibility, indicative, irrealis, non-factual, obvious, potential, realis, subjunctive	48
Deontic	deontic, desiderative, evitative, hortative, imperative, intentional, jussive, necessitative, optative, prohibitive	63
Conditional	conditional	7
Interrogative	polar, content	5
Expressive	exclamation, honorifics, stance	5
Evidential	auditive, implicative, quotative, viewpoint	4

Table 4.9: Annotated mood features. Evidentials and interrogatives have been subsumed into this category even if they might not be defined as mood proper.

Many languages of the sample exhibit quite different mood and modality features. The classification of the categories is based on how the descriptions in the grammars corresponded to the comparative concepts of epistemic, deontic, conditional, interrogative, expressive, and evidential. Some of the languages make further specifications for the features listed in Table 4.9. The category ‘expressive’ contains interjections as well as honorifics in Korean, which can be further distinguished into polite, blunt and intimate; ‘stance’ in Sheko (indirect/direct stance towards the utterance) has been also classified under the main type ‘expressive.’ Categories that refer to the source of information of the statement (evidentials) have been included in some languages (auditive in Ulwa, implicative and viewpoint in Sheko and quotative in Mian). Following the chart in Figure 4.2, modal morphs that were too specific (such as adverbials or affixes with specific adverbial meaning) were not annotated as inflectional.

4.4.1.3.3 Aspect

The selected aspect categories and features are presented in Table 4.10.

Categories	Features	Count (languages)
Bounded	anterior, completive, inchoative, perfect, perfective, punctual, resultative, telic	44
Unbounded	atelic, conative, continuative, durative, imminent, impeditive, imperfective, in-completive, partitive, simultaneous, sta-tive	48
Multiple events	habitual, sequential, serial	12
Other	Neutral, Neuter, Non-progressive	3

Table 4.10: Annotated aspect features.

There were many aspectual categories across the languages. As with mood, I grouped them into larger categories that were not expressed as features in the languages, using bounded, unbounded and multiple events as comparative concepts. Since some languages have a fine-grained distinction between aspects (such as perfective, perfect, completive, resultative in Bangime), the most common distinction 'perfective/imperfective' was not useful for naming the larger categories. Morphs in the category 'other' have either a bounded or unbounded interpretation or can co-occur with either aspect. According to Robero-Mendez (1992: 312) 'Neutral' describes an aspect that is "neutral with respect to viewpoint aspect"; In Navajo, 'neuter verbs' relate to verbs that are used as property concepts such as 'big', 'long', 'wide', and 'far' (Young, Morgan & Midgette 1992: 805) and can be conjugated with either perfective or imperfective morphology to express this function. For 'non-progressive' in Pilagá, see

Vidal (2001: 271–272). Many aspectual distinctions were therefore annotated as inflectional because they are part of other paradigms that express common TAMIP features (inclusion after phase 4 in Figure 4.2). In contrast, lexical aspectual features were annotated as non-inflectional because they are not part of paradigms of more general aspectual distinctions.

4.4.1.3.4 Indexation

The selected indexation categories and features are presented in Table 4.11.

Category	Features	Count (languages)
Person	first, second, third, specific, impersonal, indefinite, areal	35
Number	singular, dual, paucal, plural	35
Gender/Class	feminine, feminine 1, feminine 2, non-feminine, masculine, neuter, human, non-human, long object, bundle-like object, covering object, residue class, sibling, number-or morpheme named gender distinctions.	16
Clusivity	inclusive, exclusive	10
Honorifics	formal, informal, polite, blunt, familiar, intimate, plain, fourth person	5

Table 4.11: Annotated indexation features.

Annotating indexation was quite straightforward since the categories and features are very common across languages of the world (as such, most of them were included after phase 1 of the flowchart in Figure 4.2). What is not shown in Table 4.11 is that these features encode specific participants. The participants can have several semantic roles (such as actor, undergoer, recipient, addressee, etc.). In some cases, the morphs co-express properties of the event itself: in Ket, *b*- marks, besides third person inanimate, instrumental applicative and involuntary causative events, among other functions (Vajda 2004: 67); in other cases, plurality morphs also express pluractionality. I only annotated those morphs as inflectional that function to index participants. Pluractional affixes were annotated as non-inflectional if they do not also express plurality of participants in some constructions. Likewise, I annotated honorifics as inflectional that refer to participants (referents and speech-act-participants such as addressee). Honorific categories can furthermore appear associated with different vocabulary, or modifiers of speech events (such as 'blunt'), in which case they were also annotated as a modal category. Further specifications referring to property of participants have been included, such as gender, class, animacy distinctions or classifiers. Classifiers that are associated with specific events have been classified as derivational/lexical morphs and not annotated as inflectional, as was done with the Navajo thematic morph *n*- (see above under 4.4.1.2.)

4.4.1.3.5 Polarity

Finally, polarity values (positive, negative) have been annotated as inflectional. However, the sample did not exhibit languages that mark positive statements as such. Because of the universality of negation and its frequent use, negation was always annotated as inflectional. However, I only included negation markers that were listed as part of the verbal complex in grammatical descriptions. Proceeding like this enabled to capture more grammaticalized instances of negation and exclude free words that might not be as frequently used.

The number of positions with negative morphs appearing in verbal complexes is presented in Table 4.12.

Category	**Value**	**Count (positions)**
Polarity	**negative**	41

Table 4.12: Annotated polarity features

In conclusion, the different inflectional features that were annotated differ from language to language, but they were annotated as inflectional based on concrete criteria. It should be emphasized that the definition of inflection provided here is based on semantic generality and frequency and not on formal criteria.

4.4.2 Establishing morph positions

In 4.4.1, inflectional categories were determined based on semantic generality and their potential to appear in paradigms. A 'neutral ground' was established that should reflect similar high generality and similar high frequency of morphs. In this section, the method for the comparison of the different inflectional and non-inflectional categories is explained. I decided that outlining morphs on a spreadsheet according to their position in relation to other morphs would be the best solution to annotate for type (inflection/non-inflection), conditioning (LCI, GCI, UCI), prosodic prominence and obligatoriness. Thus, verbal templates for each language of the sample have been constructed.

According to Good (2011), there are morphosyntactic, morphophonological and syntactic templates. A morphosyntactic template refers to "analyses where the linear realization of components of a morphological construction is described in terms of stipulated constraints on elements characterized in terms of morphosyntactic or morphosemantic categories like *agreement* or *tense* affix" and are analyzed as "having position class" (p. 735). Inkelas (1993: 560, cited in Good 2011: 745) defines morphosyntactic templates as "morphemes or morpheme classes [...] organized into a total linear ordering that has no apparent connection to syntactic, semantic, or even phonological representation". In other words, morphosyntactic templates are stipulated when the order of morphemes does not conform to common cross-linguistic orders, such as derivational affixes occurring closer to the stem and inflectional affixes at the periphery of words (Bybee 1985).

A morphophonological template refers to "analyses where the linear realization of the components of a morphological construction is described in terms of stipulated constraints involving phonological categories" (p. 734). This template differs from morphosyntactic templates in that the features and order of consonants and vowels is relevant to interpret the morphological structure. According to Good (2011: 734), morphophonological templates are helpful to understand the morphology of Semitic languages, as well as of Sierra Miwok.

Finally, a syntactic template refers to "analyses where the linear realization of the components of a syntactic construction are described in terms of stipulated constraints on elements characterized in syntactic or semantic terms like *subject phrase* or *pronoun*" (737). Good (2011) notices that some analyses of German employ syntactic templates by dividing the linear order of words into *first position, second position, middlefield, verb cluster and postverbal field* (see Kathol 2000). As with morphosyntactic templates, the positions of categories is not always predictable from the function of these categories (verb, subject, object etc.).

The templates constructed for the samples in this dissertation do not follow the theoretical presuppositions as carried out by Good (2011), but are, like ‘inflection’, defined in relation to the goal of comparing the different variables. As a result, the templates used in this study comprise morphophonological, morphosyntactic and syntactic characteristics. However, the primary division was along morphosyntactic boundaries. That is, positions were defined as the mutual exclusion of specific morphs in linear relation to others. I implemented morphophonological templates only for stems with alternating segments, so that the semantic association of these segments could be captured below the morphological structure. Morphophonological templates are especially useful in Semitic languages where consonants and vowels have a different linguistic function (Good 2011: 734). In addition, positions were also established based on syntactic categories. In some cases, specific grammatical morphs could be separated from the verbal stem by entire phrases (such as objects). In the few instances where this was the case (Supyire, Bangime), I established a single ‘noun phrase’ position and annotated it as non-inflectional (compare noun incorporation in other languages). This noun phrase position was not further sub-divided into its morphological parts. The following subsections explain the criteria applied in determining the order and the grouping of morphs in positions.

4.4.2.1 Determining order

Highly general and frequent morphs can be individual words without occurring in a specific morphological position of verbs. As such, the distinction between ‘affixes’ and ‘clitics’ or 'word' was of secondary interest in this study. The reason why verbs are targeted is because they are the most frequently occurring part of speech with a wide array of inflectional morphology. Clitics or words that express the same verbal categories but are not bound to verbs are therefore equally of interest.

However, the syntactic variability of clitics posed challenges for translating them into specific templatic positions. A common phenomenon are so-called 'second position clitics.' These clitics usually express verbal categories but occur after the first constituent in the sentence/clause. As such, these clitics might be preceded (2a) and followed by non-verbal categories (2b).

(2) Second position clitics in Tohono O'odham (Zepeda 1983)

a.

[Hegam	O'odham]=**'o**	[cicpkan]
[DEM	Papago.person]=**3.IMPF**	[pl.work]

'Those Papago persons are/were working.' (p. 10)

b.

[Añi]=**añ**	[s-hottam]	[cipkan]
[1sg]=**1sg.IMPF**	[STAT-be.quick]	[sg.work]

'I am/was working quickly.' (p.20)

c.

[Cicpkan]=**'o**	[hegam	'O'odham].
[pl.work]=**3.IMPF**	[DEM	Papago.person]

'Those Papago persons are/were working.' (p. 10)

In order to express the position of these clitics in verbal templates, one could choose to annotate them as either preceding (2a, b) or following verb stems (2c). For this dissertation, only one order variant was chosen, namely the construction where verbs could be placed in first position and clitics followed verbs (2b), in order to avoid annotating intervening syntactic phrases as in (2b).

In the cases where verbs did not occur in first position, and inflectional categories were not bound to verbs, an extra slot for potential phrases was inserted in the template and treated as a non-inflectional position. For example, in Supyire, preverbal particles expressing inflectional categories (tense, mood, aspect, polarity) can be followed by a direct object:

(3) Inflectional morphs separated from verbs in Supyire (3) (Carlson 1994: 310)

a.

Na	wìì
1sg.NONDEC	look.at

'Look at me.'

b.

Ta	[ǹté	kyaàre]	kwùùn.
IMP.IMPF	[this	meat.DEF]	cut.IMPF

'Cut this meat.'

By allowing a syntactic expression to be included in templatic representations, some of the languages that are considered isolating or analytic did not differ from the ones that are polysynthetic in the spreadsheet. Whether this has theoretical implication for the typology of words can be investigated in further studies.

Yet other types of morphs represent challenges for establishing linearization relations between morphs. These are the morphs that can switch their positions depending on grammatical categories. For example, the personal pronouns in French can appear as enclitics in imperative clauses (*Donne=**le**=**moi**!* 'Give it to me'), but as proclitics in declarative clauses (***tu***=***me***=***le***=*donnes* 'You give it to me'). This case is different from second position clitics where verbs can occur in different positions, and the position of clitics is fixed. Here, the question arises which linearization should be preferred for the outline in the spreadsheet.

One option would be to only select specific constructions (as done with second position clitics constructions). Choosing the most frequent construction can be a solution that complies with the goal of this study. In discourse, declarative clauses are more frequent than imperative clauses, and have a higher type frequency. This high type frequency has more impact on the structural evolution of clauses as they might be more/earlier entrenched in the mind of speakers. Thus, if one chooses only declarative constructions in French, the clitic placement is more fixed. Ashby (1977: 16) defines three templates for the 'verbal group' in French: finite, infinitive, and imperative. Thus, focusing on finite verbs (encompassing indicative and some interrogative constructions) would yield a unitary template for French.

A second solution is to annotate the category of the morph in all positions where the morph appears. Another example from French is the linear behavior of the negative morph 'pas' /pa/. 'Pas' can occur before the object clitics in desiderative sentences, such as 'Je ne veux **pas** m'asseoir' (I don't want to sit down), after the finite verb in statements such as 'Je ne m'assieds **pas'** (I don't sit down) or between auxiliary and participle, such as in 'je ne me suis **pas** assis' (I didn't sit down). Here, it is harder to deduce from grammatical descriptions which construction is more frequent. One would have to either randomly choose a position that seems more intuitive, or annotate separate positions for the same category, like in the following linearization:

'SUBJ-ne-AUX(vouloir)-pas-OBJ-AUX(être)-pas-VERB(finite/infinitive/participle)-pas'

This solution might entail semantic redundancy; however, it improves the alignment with the phonological schema of verbal constructions and allows more specific predictions of the relation of a position with prosodic prominence. In French, 'pas' is only prosodically prominent when it occurs after lexical verbs (i.e., not after auxiliaries), since stress in French falls on the last syllable of the phonological phrase.[13] When annotating multiple positions for 'pas', it will be captured that 'pas' is always stressed in certain constructions, and unstressed in others.

A third solution would be to annotate only one position of the morph but take into consideration the properties of the morph when it occurs in another position. With regards to French 'pas', one could annotate one position where 'pas' occurs and choose the label 'possible' for stress, instead of annotating two or three positions and choosing 'yes', 'no' and 'no'. This solution would treat the morph 'pas' as one but also hide the information about variation in positioning. Because this dissertation takes into account the phonological dimension of schemas, it has been decided to choose the second solution for annotating morphs that vary across

[13] The question of whether French exhibits stress as opposed to another property that marks prominence in phonological phrases is an old topic (Verluyten, 1984; DiCristo, 1999; Lacheret-Dujour and Beaugendre, 1999). Phonological phrases in French are most prominent in their last non-schwa syllable, and this is considered as evidence of stress in some analyses (e.g., Temperley & Temperley 2013). However, Peperkamp, Dupoux, & Sebastián-Gallés (1999) found that French speakers have difficulties in distinguishing two words that differ only as far as the location of stress is concerned, and that French stress is not contrastive. For this dissertation I will use 'stress' with regards to prosodic prominence marking in French but acknowledge that it is not a feature of morphological words and not lexically contrastive.

very frequent constructions. The first solution (focusing on few constructions to reduce templatic variability) cannot be straightforwardly carried out. Thus, clitics that change the position in constructions of similar type frequency should be annotated in several positions. This ensures that the information on the alignment with prosodic prominence is more detailed.

4.4.2.2 Grouping morphs

As mentioned above, morphs were aligned into templatic positions. The segmentation into positions made it easier to account for the effect that paradigms have on the mind of speakers (in contrast to single morphs). Since only inflectional positions were selected for the annotation of morphological conditioning, the question arose what to do when one position exhibits both inflectional and non-inflectional morphs. To give an example, the 13th suffix position in Koasati (Muskogee) contains the following morphs:

(4) Morphs of suffix position 13 in Koasati (Muskogee) (Kimball 1991: 208-212)

=*:s(a)* (Past I)

=*t(i)* (Past II)

=*t(o)* (Past III)

=*k(i)* (Past IV)

=*:fó:k(a)* (when; while)

=*V́n(na)* (negative imperative)

Every morph in this position is phonologically conditioned (the variants being *t/ti*, *t/to* etc.). According to the selection process, only the past tense markers (Past I, II, III, IV) and negative imperative are classified as inflectional, not the conjunction-like clitic *=:fó:k(a)*. The solution applied here was to treat the entire position as inflectional, and to ignore lexical or derivational/adverbial morphs in it. Position 13 in Koasati was thus annotated as UCI (unconditioned inflection; encompassing also phonologically conditioned allomorphs).[14] In consequence, only the structural values (prosodic prominence, obligatoriness) of these inflectional morphs were annotated for the entire position. This grouping and filtering of morphs is advantageous in the following ways.

First, it facilitates the selection process between inflectional and derivational morphs. Once one inflectional morph is annotated as such, the other morphs in the same position can be ignored for the selection process. This way, one does not need to split positions and analyze every morph. Positions are annotated as either inflectional or non-inflectional. In fact, due to the paradigmatic nature of inflection, there were not many cases where derivational morphs co-occupy a position with other inflectional morphemes.

Second, highlighting inflectional categories and de-emphasizing non-inflectional (lexical, derivational, adverbial) categories resulted in a more balanced distribution between these two categories. Grouping together inflectional morphs avoids coding similar features (1sg.past, 2sg.past, 3sg.past) multiple times. Derivational or adverbial morphs are more 'spread out' across templates, that is, they are more agglutinative and are less likely to form paradigms. For example, Ayutla Mixe has 11 prefix positions, but only position -11 contains inflectional

[14] The different numbers for past and their similar shape might suggest that these morphs are cumulative; however, the description of the past tense morphs in Kimball (1991: 208) does not allow such an interpretation. The additional meanings of past do not seem like features of a grammatical paradigm distinct from past, but part of the meaning of past (Past I= past from the viewpoint of a hearer; Past II very recent past; Past III = past from several hours to several years ago; IV: past from many years ago). In Section 4.5.2.1, I explain more clearly why this type of inflection was not included in the GCI category.

affixes. The number of morphs in this position is however higher than the number of morphs in the non-inflectional positions. Position -11 contains 10 morphs, and each of them mark participant scenarios (encoding person of both agent and patient), whereas each of the remaining 10 positions contain only one adverbial/valence-changing prefix, such as P-9 *ës-* ‘motion cum purpose’ (P-9), *ak-* ‘causative’ (P-6) or *ta-* ‘instrumental applicative’ (P-5). By ignoring occasional derivational/adverbial affixes in inflectional paradigms, the number of tokens that were relevant for the investigation increased, and the number of tokens that were not as relevant for the investigation was minimized.

Third, basing the spreadsheet on positions and not on single morphs complies with the variables for structural stability presented in Chapter III. Stress patterns are less dependent on specific morphs (although some morphs can attract stress), but rather on the prosodic position of the word. This means that if a word is always stressed on the last syllable of the word, and the last morphological position contains an entire paradigm of morphs, every morph of this paradigm is stressed. Annotating the stress pattern for each morph would not capture the prosodic effect on the entire position. Furthermore, the positional lumping is indispensable with regards to the second stabilizing variable, obligatoriness. Treating each morph individually would mean that no morph would receive the feature ‘obligatory’, since a specific morph is never obligatory for all (schematic) verbal constructions. Instead, obligatoriness refers to the entire paradigm in a specific position and should account for the high type frequency of its morphs. Lumping morphs into one position enables judging whether the entire paradigm (which in most cases aligns with a specific position) is obligatory.

4.4.2.3 Splitting positions

Sometimes morphs from the same position were annotated in different positions of the Excel template. In order to increase informativity, positions were 'split' when inflectional morphs from different grammatical categories differed in the value of one variable, and labeled as 'sub-positions' such as 1.1, 1.2, 1.3 etc. For example, in Kiowa, the first suffixal position (P1) contains morphs from three categories: valence (transitive/stative/causative morphs), perfective (4 allomorphs) and imperfective (6 allomorphs). Following the procedure outlined above, these categories would be grouped together, and the first category (valence) ignored. However, imperfective and perfective allomorphs differ in the tone pattern. Perfective allomorphs always bear a high tone *(-ɔ́, -iá(y), -é, -Ø*), whereas imperfective allomorphs have a low or falling tone *(Cɔ̀/-gù, -(m)ià, -p, -î, mà*) (Watkins 1980: 206; 202). According to the criteria further laid out in 4.6.1.5, the prosodically prominent tones in Kiowa are the high and falling tone. Thus, these two allomorph sets differ in prosodic prominence. Furthermore, perfective and imperfective allomorphs also differ in conditioning: perfective allomorphs are lexically conditioned, imperfective allomorphs are grammatically conditioned.

Without this splitting, the more precise information between conditioning type and prosodic prominence would not be captured. The coarse alternative would be to annotate this position as LCI (since GCI morphs are ignored when the position contains LCI morphs as laid out further in Section 4.5.1), and to assign the value 'possible' for prominence (instead of 'yes' and 'possible' for LCI and GCI respectively). As a result of allocating these two allomorph sets into two positions, the derivational morphs of P1 in Kiowa were annotated in yet another position (in order to avoid having to choose which inflectional position the derivational morphs would be assigned to), and thus, P1 was split into the three positions P1.1, P1.2 and P1.3.

However, if the morphs belonged to the same paradigm/category but showed different values for either prosodic prominence or obligatoriness, the position was not split. Thus, positions were only split if the resulting position shared the same semantics (e.g., an aspectual position could be split into imperfective and perfective allomorphs, but an indexical position could not be split into one position containing first person and another position containing second and third person), and if conditioning and at least one value of one stabilizing variable (prosodic prominence or obligatoriness) was different after the splitting. In the template, these positions are still indexed with the same number.

By splitting a position in semantically defined sub-positions, one would expect the value for obligatoriness to change. As stated further in 4.6.2.3, obligatoriness was defined in relation to the filling of one templatic position. One could imagine a scenario in which a position is always filled, for example by tense suffixes, but these suffixes have different conditioning values (LCI for past, UCI for future). By splitting the position into multiple positions, the new positions (e.g., past, future) are no longer obligatory. In this case, I annotated for each position the value that the position had before being split. In practice, I only split one obligatory position, namely in Yurok. Yurok has two sets of obligatory subject affixes (o- and e-class), but only the o-class affixes can be prosodically prominent, if the stem of the verb is monosyllabic (Robins 1958: 34). Therefore, the suffix position P2 has been split and both positions were annotated as obligatory.

Stems were often split. There were phonological reasons for this choice. Several languages contained stems where vowels or consonants encoded grammatical function (stem alternation). For example, the shape of stems in Yurok can be further analyzed as having an initial, medial and final component (Blevins 2005). The initial component encodes lexical information, the medial component lexical and/or derivational information and the final component lexical, adverbial, valency and mood (imperative). Because lexically conditioned inflection is often expressed via stem alternation, splitting up stems into inflectional and lexical components yielded more data that could be compared. Splitting up stems did not result in fewer obligatory positions, since stem alternation usually occupies segments that are obligatory in the phonological make-up of the stem.

In conclusion, both lumping and splitting positions had the goal of increasing the number of inflectional tokens to be compared with the values of prosodic prominence and obligatoriness and to yield more significant results in this comparison. The next section explains how the complexity variables (LCI, GCI, UCI) were established and the values determined.

4.5 Complexity variables

Since this dissertation investigates the effects of structural environments on the distribution of complex morphology in contrast to less complex morphology, the operationalization of the parameter 'complexity of conditioned inflection' is needed. The variables must reflect a degree in this parameter, namely between 'more complex' and 'less complex.' One suggestion brought forward in Chapter II is to select qualitative distinctions that, in relation to one another, inherit this gradience. This was based on informational-theoretic insights but more importantly, on the consistency of psycho- and neurolinguistic studies showing increase in processing effort. After narrowing down the scope of inflection by focusing on inflection of verbal constructions (Section 4.4), the distinctions LCI, GCI and UCI need to be established. For the creation of the main sample, the existence of LCI in verbal inflection was the only criterion. This means that I did not control for a certain number of GCI or UCI in the LCI sample. Instead, the gathering of GCI and UCI morphs was a result of including LCI verbal templates into the sample. Luckily, these languages provided a balanced number of GCI and UCI morphs in comparison to LCI morphs (see Table 4.13 in 4.5.1.2). This section explains how the inflectional positions were classified into either LCI, GCI and UCI.

4.5.1 Determination of LCI positions

The purpose of finding lexically conditioned inflection was to establish maximal complexity. As laid out in Section 4.4, only inflectional positions were coded for type of conditioning. This does not mean that derivational morphs cannot be lexically conditioned. Indeed, lexicality (the case where morphs attach to a specific group of stems that cannot be semantically or phonologically defined) increases the more derivational a grammatical category is, because derivational morphs are more relevant to the meaning of the stem (Bybee 1985: 15ff.). Section 4.5.1.1 explains the criteria according to which LCI morphs have been annotated as such.

4.5.1.1 Criteria

The effects mentioned in experimental studies concerning irregular vs. regular derivation are not as conclusive as with inflectional morphs (Leminen et al. 2019). It is more likely that words with derivational morphology are memorized through declarative knowledge, no matter whether the derivational morphs are lexically conditioned. Words with inflectional morphs show a clearer dichotomy in the effect on speakers when confronted with LCI vs. GCI/UCI, because of the discrepancy that consists in these morphs being semantically quite general and frequent (they occur in many words) but restricted to specific lexemes. As such, I only included LCI morphs where this discrepancy was most evident, namely when the conditioned and conditioning contexts are clearly defined. This means that suppletion, where it is not clear whether the lexicon conditions inflection or inflection conditions the lexicon, is not a good candidate for high complexity. Likewise, single irregulars, where the lexical restriction of an inflectional allomorph is conditioned by a single word, are less likely to challenge the mind of speakers, also because of their high token frequency; inflection in high token frequency words is more likely to be processed by declarative knowledge together with the word including their morphological structure (Bybee 1995: 433-435). In contrast, low token frequency of words is more likely to challenge speakers in choosing the correct LCI morph, and they have to apply both declarative and procedural knowledge. This is the case with large lexically conditioned paradigms (‘lexical classes’). The greater the paradigm of inflectional morphs, the lower the frequency of a specific morph in that paradigm, and the more likely speakers would apply procedural knowledge to these.

Because of the reasons mentioned above, single irregulars and suppletive paradigms were excluded from annotation, and instead only affixes which belonged to a paradigm that was used for more than one word (including stem alternation) were chosen. The minimal paradigm consisted of two morphs for one inflectional category. This would include for example English irregular classes containing forms such as ‘drink/drank’ and ‘ring/rang’ (both form their present tense with the stem vowel /ɪ/ and past tense with the stem vowel /æ/).

In order to maximize the tokens of LCI (since LCI is rarer than other types of inflection), positions were annotated as LCI even if not all morphs in the paradigm were lexically conditioned. For example, in Skou, not all verbs show lexically conditioned stem alternation (Donohue 2004: 228). Still, the stem position has been annotated as LCI, while ignoring those stem morphs that do not exhibit LCI. In Yurok, subject indexation is lexically conditioned except for the 1st person plural, which is -*oh* in both word classes. This morph was therefore ignored for the subsequent annotation of LCI with the other variables: that is, prosodic prominence and obligatoriness values have been assigned only for LCI in the paradigm. Furthermore, in order to uphold a tripartite classification, the most complex conditioning type of morphs has been annotated. Thus, positions that exhibit LCI morphs that were also grammatically or phonologically conditioned were annotated as LCI positions. For example, in Navajo, tense, mood, aspect and indexation are marked in P-2. The fusional status of these morphs means that they are grammatically conditioned. However, these morphs belong to larger sets whose choice is conditioned by the stem and other lexical affixes. As such, position -2 in Navajo was annotated as LCI.

It was not always straightforward to determine LCI positions. In few cases, authors emphasized that the specific morphs or paradigms were lexically conditioned but used other descriptions instead. Lexical conditioning was inferred when authors mentioned that paradigms were not always predictable for every verb, or that one paradigm/morph was used for few verbs. Some examples are given below:

> "The reason why [...] plural number of the subject is indicated through the object number markers, rather than by subject number markers, has no explanation to my knowledge. One possibility is that this plural marking system may be the residue of a no longer productive absolutive agreement system, or rather the opposite, it could be an emergent absolutive system, and for either reason sometimes functions to indicate the plural number of the only argument of an intransitive verb (Vidal 2001: 165, about number suffixes in Pilagá)."

> "Most dynamic verb stems require subjective prefixes, except when the stative suffix is added, while others require objective prefixes." (Abbott 2000: 38 about subject-object prefixes in Oneida)

> “The phonological shape of modified roots vis-à-vis basic roots is determined in part by a categorisation of all verbs as belonging to the class of either weak or strong verbs, and partly on the basis of the phonological shape of the root itself (...) There is an element of unpredictability in the subclassification of verbs in Ura into these two groupings.” (Crowley 1999: 151 about the stem alternations in Ura).

These descriptions show that authors prefer words like ‘idiosyncratic’, ‘arbitrary’ ‘non-predictable’, instead of ‘lexically conditioned.’ Non-predictability does however not always mean lexical conditioning. In other cases, it relates to free variations that speakers use, which make grammatical rules non-predictable as well. These cases of free variation have not been annotated as LCI and there are good reasons to assume that free variation does not challenge speaker’s mind in the same way as LCI does: if morphs are fully arbitrary, speakers are free to choose between morphs without being sanctioned by others or themselves.

Finally, paradigms were annotated as LCI when the lexicon conditioned either form or meaning. For example, in Ket, prefix position 6 (marked in bold) contains morphs that can refer to patient or agent, depending on the verb (‘go’, ‘sell off’, ‘bring away’).

(5) Lexically conditioned semantic role assignment of Position 6 morph *kú-* in Ket

a. kú-ɣ-a-tn

2.A-TH-PR-go

‘**You** go.’ (Vajda 2004, p. 51)

b. d-**kú**-k-dì-qa

1.A-**2sg.P**-TH-1sg.A-sell

‘I sell **you** off.’ (Georg 2007: 199)

c. k-**kú**-k-dì-qos

2.A-**2sg.A**-TH-1sg.O-bring

‘**You** bring me away.’ (Georg 2007: 197)

4.5.1.2 Distribution

By applying these criteria to the annotation of inflectional morphs, the following distribution across positions of the LCI sample was obtained (Table 4.13).

	Preceding	**Stem**	**Following**	**Total**
LCI	20	28	32	80
GCI	37	1	37	75
UCI	28	1	32	61
Non-Inflection	62	13	36	111
Total	147	43	137	327

Table 4.13: Number of position types (preceding, stem, following) of the LCI language sample. This includes split positions, as included according to criteria in 4.4.2.3.

The 31 languages of the LCI sample yielded 327 morphological (sub-)positions (including stems, affixes and clitics). Out of the 327 positions, 216 contained inflectional morphs and 111 merely non-inflectional morphs. The LCI sample also shows enough inflectional positions with a different conditioning type. As such, the sample lends itself to investigating correlations of conditioning types within one language type (languages with LCI). As one can see in Table 4.13 and Figure 4.3, non-inflectional positions tend to precede the stem morphs (as prefixes or proclitics), whereas inflectional positions precede or follow the stem equally often. Most notably, LCI positions very frequently are stems or segments of stems. On a closer examination, LCI tends to also be closer to the stem than GCI and UCI, and GCI more so than UCI. Figure 4.3 shows the distribution of conditioning types across positions of LCI languages. The distribution reflects most prominently the rate of positions relative to stems: only a few languages have more than 5 prefix positions and more than 5 suffix positions. The non-inflectional positions are commonly associated with proximity to the stem (in the case of derivational affixes) but also appear at the outmost periphery in large templates, as adverbial clitics.

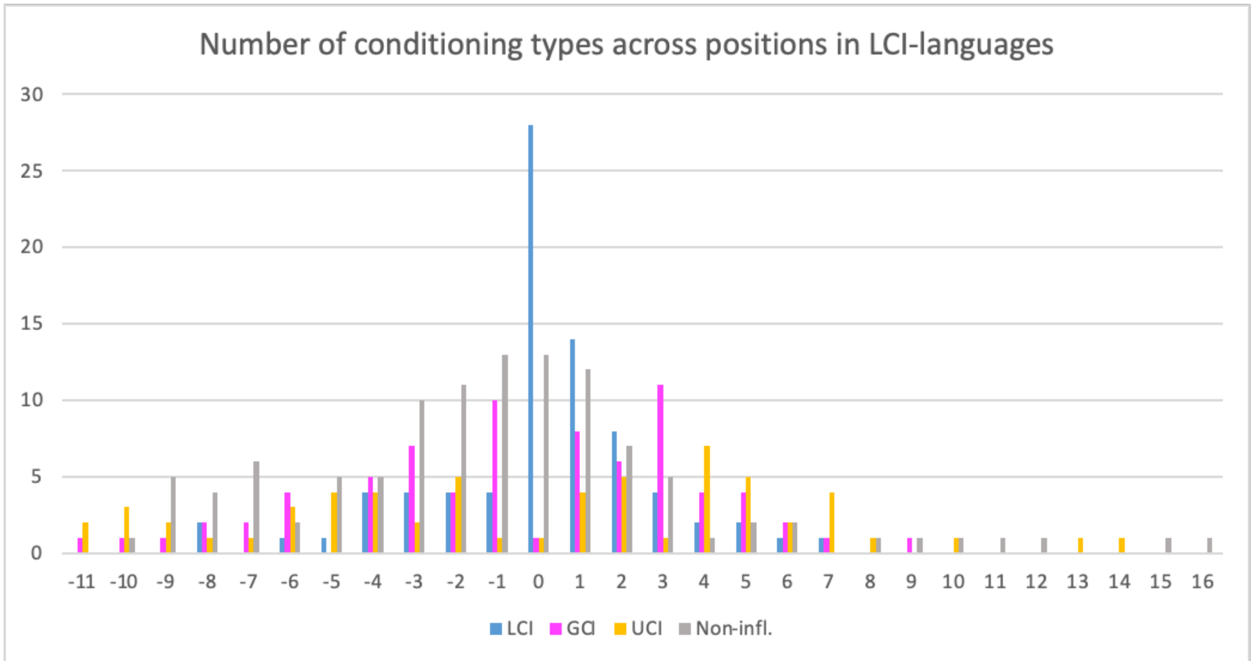


Figure 4.3: Distribution of conditioning types across prefix (-11 – -1) and suffix (1 – 16) positions in LCI languages.

Because of the likelihood for LCI to be expressed by stem alternation, the discussion of LCI will require a separate examination of stem vs. non-stem morphs when contrasting it to prosodic prominence and obligatoriness. As noticed in Chapter III, stem morphs bear properties that are different from non-stem morphs, such as more semantic prominence and morphological obligatoriness. That is, because stems are almost always obligatory, and almost never contain non-lexically conditioned inflection, non-stem LCI might be separately investigated with regards to obligatoriness. Likewise, stems themselves might also entail a different relation to prosodic prominence as they are more likely to attract stress (cf. Beckman 1998; Revithiadou 1998). Nevertheless, obligatoriness and stress might still have an impact the distribution of LCI in stems. LCI can be the cause (LCI is present because stems tend to create LCI) or effect (LCI is present because stems tend to preserve LCI, by means of being prominent and obligatory) for the association of LCI and stems. Specific properties of LCI morphs (especially stems) might be directly correlated to the *creation*, and not *retention* of LCI. Therefore, it is more plausible to distinguish cause and effect for positions where all types of conditioning could be equally present. That is, if prominence and obligatoriness correlate gradually in relation to LCI, GCI and UCI, it is more likely that the reason for the distribution is the perpetuating effect of prosodic prominence and obligatoriness, and not of the position type (stem vs. affix). These issues are further discussed in Chapter V.

4.5.2 Determination of GCI positions

The purpose of determining GCI positions was to establish medium-degree complexity. According to the investigation in Chapter II, GCI is argued to be less complex than LCI but more so than UCI. Morphs conditioned by grammatical categories, grammatical morphs/positions as well as cumulative morphs have been classified as GCI morphs. Following the flowchart in Section 4.4.1.2, grammatically conditioned *derivational* morphs (for example, tense-governed choice of valency) have been excluded. Furthermore, like with LCI morphs, some GCI morphs have been excluded from the annotation.

4.5.2.1 Criteria

First, cumulative morphs were excluded where one feature that was cumulatively expressed did not pertain to a paradigm, that is, did not exhibit another corresponding feature. Examples are given below:

(6) Semelai (Kruspe 2004)

ma-: irrealis/potential, agentive (p. 161)

-ɲá: imperative, transitive. (p. 417)

These two morphs comprise several meanings in Semelai, and as such can be seen as cumulative. However, they have been annotated as UCI. The reason for doing so is that while *ma-* and *-ɲá* combine two grammatical features from different functional domains (valency and mood), there are no other allomorphs. This means, there is no morph for 'realis/potential.passive' or 'imperative.intransitive'. In contrast, K'ichee' has imperative allomorphs, both for transitive and intransitive contexts:

(7) K'ichee' (Mondloch 2013)

-*oq:* imperative in final position, intransitive (p. 41)

-*a:* imperative, non-final position (p. 41)

-*oq*/*qa*: imperative passive for monosyllabic stems (p. 122)

-*a'*/*o'*/*u'*: imperative for transitive monosyllabics (p. 115)

The annotation of the morphs in (7) as GCI results from the interaction of valency and mood; phonological sub-specifications (finality, syllable type) are ignored with regards to the GCI category.

Second, morphs that are used in different meanings that are quite close or whose usage is not defined by a specific context have been excluded. For example, the suffix -*(y)AcAX* in Azari (Lee 1996: 47-48;51) denotes 'intentionality' and 'future tense', two meanings that are closely related. Because their semantic alternatives are similar, it is questionable that speakers would show effort regarding the memorization of these two distinct semantic features. When the fused meanings can be generalized by an overarching meaning (intentionality entails futurity) or closely related meanings, these meanings might be conceptualized as one, or one meaning might even facilitate the memorization of the other.

Third, cumulative morphs whose occurrence is exceptionally high, such as person-number-clusivity combinations, should not be annotated as GCI but as UCI. Person-number-clusivity combinations have a high token frequency, which entails a unitary memorization of the features. This type of cumulation might be even *less* complex than agglutination, because cumulation of person, number and gender might co-facilitate the retrieval of participants. Consequently, indexation conditioned by other indexation categories has been annotated as GCI only when they belonged to two paradigms encoding different participants, like agent-patient combinations. Participant-participant cumulative morphs have a lower token frequency and higher type frequency, which means that speakers are more likely to process the categories as distinct and less likely to treat them as a unitary entity. Likewise, morphs that co-index features of different participants (e.g., subject/object) or mark participant-participant relations (such as different morphs for direct and inverse scenarios conditioned by an indexical hierarchy, see Zúñiga 2006) have been annotated as GCI since the combination of properties from two participants determine the choice of morphs.

Fourth, indexation of gender or other attributes of referents was encoded as GCI when it was not predictable from the semantics of the referent. Gender can be predictable in relation to the nouns it indexes (e.g., a feminine index for referents that are female, animate for humans), but in most cases, gender is unpredictable from the semantics of the referents. In the first case, one could argue that the index morph is not grammatically conditioned but is an inherent property of indexation (indexation is always dependent on properties of referents). In the second case, one could argue that for non-predictable gender assignment, gender indexation is lexically conditioned (namely by specific nouns that are indexed through a specific gender morph). However, it is questionable whether lexical classes of nouns are more complex than lexical classes of verbs. After all, the specification of gender classes is functional in the sense that it facilitates reference tracking (Trudgill 1999; Comrie 1988; Croft 1994), whereas the function of lexically-conditioned verbal gender indexation is less evident. I decided to annotate gender indexation as GCI when the gender of referents was unpredictable, as this would account for their lower complexity in contrast to morphs conditioned lexically by the verbal construction. Not annotating gender morphs as LCI also allows us to annotate them as LCI when there are more morphs that are conditioned by the verbal lexicon. For example, in Ket, there are several gender allomorphs, depending on which verb they are attached to:

(8) Lexically conditioned gender indexation in Ket

a. **ú**-ɣ-ȁ-tn

3.F-TH-PR-go

'She is going.' (Vajda 2004: 50)

b. **də́**-n-ò

3.F-TH-PR-go

'She died.' (Vajda 2004: 51)

In conclusion, morphs that index inherent semantic properties of referents such as discursive function, animacy or sex have been annotated as UCI, and indexation that is conditioned by a lexical class of referents have been annotated as GCI, and indexation morphs that index inherent semantics of referents or are conditioned by a lexical class of referents *and* conditioned by verbal lexemes/lexical classes have been annotated as LCI.

Fifth, so-called 'formatives' that are required for inflectional paradigms have been annotated as GCI. In this dissertation, formatives are defined as morphs that lack an individual meaning but are nonetheless needed for morphological constructions. In other words, formatives change their meaning based on the construction they appear in and are by definition grammatically conditioned. For example, in Kinyamwezi, the suffix *–a* appears (among other contexts) in Future, Immediate Past, but not (among other contexts) in Optative:

(9) Formatives in Kinyamwezi (Maganga 1992: 103–104)[15]

a. a-**kʊ**-mal-**a**

CL1-**FUT/HAB**-finish-**FORM**

'S/he will finish.'

[15] These examples were not glossed and translated in the grammar (Maganga & Schadeberg 1992). As such, the glossing and translation is provided according to my best judgment on the information provided in the description. The morphs 'FUT' 'NARR' 'OPT' and 'DIST FUT' might be also analyzed as formatives but were glossed as feature-bearing morphs in order to contrast the behavior of formatives in a few examples. The entire table can be seen in Maganga & Schadeberg (1992: 103–104).

b. a-**ka**-mal-**a**

CL1-**NARR**-finish-**FORM**

'S/he finished.'

c. a-mál-**eé**

CL1-finish.**OPT**-**FORM**

'S/he might finish.'

d. a-**laa**-mal-**eé**

CL1-**DIST.FUT**-finish-**FORM**

'S/he will finish.'

One could argue that because these morphs lack a true 'conditioned' feature, they should be treated as UCI. However, the reason for annotating them as GCI results from the form-meaning mismatch – there are alternative meaning interpretations for these morphs dependent on other morphs and constructions. Thus, they can be regarded as a type of syncretic morphs, which have been classified as grammatically conditioned morphs in Chapter II.

Sixth, so-called 'auxiliaries' which are part of verbal complexes have been annotated as GCI since they comprise more than one grammatical feature. Auxiliaries are frequent lexemes that express grammatical functions, and as such, have more in common with other grammatical morphs than lexical stems. Auxiliaries exhibit at least one phonological portion that can be associated with their stem; the grammatical features expressed by this stem is taken as the conditioned feature. Categories expressed by affixes or stem alternations in auxiliaries were annotated as the conditioning feature. The following table shows these two types of auxiliaries, in French and Ingush.

Person/ Number	**Inflected auxiliary**
1sg	vɛ
2/3sg	va
1pl	alõ
2pl	ale
3pl	võ

Table 4.14: Stem alternating inflection of the future auxiliary 'go' in French

Gender	**Inflected auxiliary**
V	vy
J	jy
D	dy
B	by

Table 4.15: Agglutinative inflection of future auxiliary 'be' in Ingush (Nichols 2011b: 228-229). The genders are named by the prefixed consonant (V-gender for *v-y*; J-gender for *j-y* etc.)

This also solved the problem of how to account for nesting morphological structure in verbal templates. No distinction was made for whether inflection in auxiliaries was fusional or concatenative. The fact that marking of future in French and Ingush entails the marking of indexation features makes future constructions more complex than when there is a dedicated morph for future. In Ingush, the future tense construction entails multiple indexation of the absolutive argument:

(10) Auxiliary agglutination in Ingush (Nichols 2011b: 357)

Handza	Mariem-az	niw	hwa-**j**-iel-ag-[**j**-y]
In.just.one.moment	Mariem-ERG	door(J)	DX-**J**-open-FUT-[**J**-be]

'Mariem will open the door in just a moment.'

Derivational auxiliaries with inflectional morphs were not annotated as inflectional since the conditioned element – the auxiliary stem – was not. However, auxiliaries that are used to form the lexical entry of the verb *and* exhibited inflectional morphs were annotated as LCI, parallel to annotating LCI for inflected stems. In Murrinh-Patha, the lexical information of verbs consists of an auxiliary and a stem, separated by inflectional morphs. Note in (11) how verbs require different auxiliaries for the features '1sg.FUT'.

(11) Lexical auxiliaries in Murrinh-Patha (Walsh 1976)

a. ***ŋiɹa***-*murk-nu*

1.sg.FUT.stand-throw.away-FUT

'I shall throw (it) away frequently.' (lit. 'I will throw while standing.') (p. 229)

b. ***ŋiɹa***-*nu*

1sg.FUT.stand-FUT

‘I will stand.’ (p. 332)

c. ***ŋu***-*pup-nu*

1sg.lie.FUT-lie-FUT

‘I will die’ (lit. ‘I will lie down while lying down.’) (p. 230)

d. ***ŋu***-*nu*

1sg.lie.FUT-FUT

‘I will lie.’ (p. 346)

The examples (11a–d) show that the first morph can act as the sole finite verb of the sentence or combine with other stems to form complex predicates. These auxiliaries do not actually express derivational information, because the stem ‘throw away’ always requires the auxiliary. The shape of the tense-person auxiliary is not predictable, and as such it is lexically conditioned by the auxiliary.

Parallel to the procedure with LCI positions, positions that contain morphs that are grammatically conditioned but additionally grammatically or phonologically conditioned were coded as GCI. Thus, morphs conditioned by multiple grammatical categories were annotated as GCI, but morphs that were conditioned by grammatical categories and the lexicon as LCI. In addition, following the criterion where only the highest complexity degree was annotated, positions with both GCI and UCI morphs were annotated as GCI.

4.5.2.2 Distribution

GCI positions comprise a heterogeneous variety of morphs – heterogeneous because the conditioning and conditioned categories can vary. In contrast, LCI morphs are always conditioned by the category 'lexeme (class).' The 94 GCI positions can be mainly divided into the ones where the conditioning-conditioned relation between categories is discernible (79 positions), and into the ones where this is not the case (cumulative morphs) (15 positions). Table 4.16 shows the distribution of the categories and positions where the conditioning-conditioned relationship is evident.

		Conditioned categories							
		Aspect	Tense	Mood	Formative	Polarity	Indexation	Tot.	# Pos.
Conditioning categories	Valency	1	2	2	1	1	11	18	19
	Aspect	4	4	2	5	7	9	31	30
	Tense	0	0	0	6	7	8	21	22
	Mood	0	0	0	5	4	12	21	26
	Polarity	1	1	2	1	0	5	10	7
	Indexation	4	2	3	0	3	6	18	16
	Syntax	1	1	3	2	4	11	22	23
	Total	11	10	12	20	26	62		
	# Positions	11	10	11	9	10	35		

Table 4.16: Number of categories and positions exhibiting these categories across GCI positions (Total 79). The number of categories and positions exhibiting these categories is greater than the total number of GCI positions, because certain positions exhibit morphs expressing and being conditioned by several categories.

Table 4.16 reveals that indexation is the category that is most often conditioned by other categories, both regarding the number of positions (35) and the number of different categories conditioning it (62). This is because indexation is often co-conditioned by other indexation categories (such as subject-object scenarios), whereas other categories are not (tense, mood, polarity). On the other hand, aspect features are the ones that most likely condition other features (30 positions, 31 different categories). Tense and mood have no specific likelihood of being either conditioning or conditioned categories. A closer look at the distribution of conditioning and conditioned categories reveals that aspect can condition all other inflectional categories, and indexation can be conditioned by all other inflectional and non-inflectional categories. Comparing this behavior to Bybee's (1985) cline where aspect is most relevant and indexation least relevant, one can say that conditioning relationships are manifestations of relevance relationships. According to Denk (2019), relevance relationships can be defined between affixes, not only between affixes and stems.

There are no GCI instances where Tense and Mood condition other tense and mood and aspect categories. Formatives are by definition only conditioned categories, never conditioning. Polarity behaves like indexation and aspect in that it conditions and is conditioned by all other categories; however, like indexation, it is more likely that polarity is conditioned by other categories than it conditions other categories. The category 'valency' and 'syntax' (when phrases or properties of other words, such as gender, condition morphs in verbs) appear only as conditioning categories since the opposite (inflectionally-conditioned valency or syntax) would not be considered an inflectional category, and therefore not annotated as GCI.

Table 4.17 shows the number of cumulative positions across the inflectional categories.

	Aspect	Tense	Mood	Polarity	Indexation
Number of positions	8	8	8	4	9

Table 4.17: Number of GCI cumulative positions exhibiting inflectional categories (Total: 15).

Table 4.17 shows that the inflectional categories are evenly distributed across cumulative positions. Polarity, due to its lower frequency of occurrence as an inflectional category, is understandably less represented in cumulative positions.

In conclusion, conditioned, conditioning and cumulative inflectional categories are well represented across the 94 GCI morphs from 36 languages from the LCI and GCI samples, which constitutes enough data to establish a separate category of GCI. Due to the heterogeneous character of GCI positions, it is expected that this category might not behave as uniformly as LCI or UCI when correlating them to their structural properties (prosodic prominence, obligatoriness).

4.5.3 Determination of UCI positions

The third type of conditioning defined for annotation is ‘unconditioned inflection’, (UCI). UCI encompasses positions with morphs that have no allomorphs or morphs that are merely phonologically conditioned (PCI). In Chapter II, it was argued that there is little evidence that PCI poses a challenge for acquisition or memorization as opposed to unconditioned morphs. Indeed, in some cases, PCI might be more efficient than having no allomorphy (e.g., Martin and Peperkamp 2020). Based on this reason, the category UCI has been determined to also include PCI allomorphs; the ‘unconditioned’ attribute refers more strictly to ‘unconditioned by grammar or lexicon.’ In the spreadsheet, I have kept track of any information regarding phonological conditioning. However, I did not further investigate phonological conditioning of allomorphs if I had to infer it from phonological rules mentioned in other parts of the grammar.

4.5.3.1 Criteria

Because UCI is defined negatively, it is a residual category. The following paragraphs describe how UCI positions were determined as such.

First, morphs were annotated as UCI when they expressed only one grammatical feature (non-cumulative except indexation combinations of person, number, predictable gender and clusivity) and could occur in the most frequent verbal construction type (i.e., finite verbs), unconditioned by any other grammatical categories or lexemes.

Second, indexation morphs that are part of grammatical relations were annotated as UCI. According to LaPolla (2006) "grammatical relations are construction-specific conventionalizations (grammaticalizations) of implicatures which arise out of repeated patterns of reference to particular types of referents." More specifically, one can say that grammatical relations are neutralizations between semantic roles of intransitive and transitive participants. The category 'Subject' is such a grammatical relation, a conventionalized category resulting from the neutralization of A (agentive transitive argument) and S (single intransitive argument) across verbs. The category 'Absolutive' is instead a neutralization of S and P (patientive transitive argument). Assuming that neutralization of roles is an efficient strategy that reduces overspecification of referents and therefore complexity for speakers, the grammatical relations and roles were not annotated as additional or conditioning features of morphs. Thus, if indexation morphs were not conditioned by any other categories, these were annotated as UCI. Likewise, morphs involved in secundative (neutralization of direct object of transitive clause & indirect object of ditransitive clause), and indirective alignment (direct object of transitive clause & direct object of ditransitive clause) were also not regarded as grammatically conditioned. On the other hand, the rarer pattern of intransitive split-alignment was considered GCI, where the choice of indexation morphs is dependent on the semantics of participants. Likewise, hierarchical alignment was encoded as GCI, where the choice of indexation morphs is dependent on the relational properties of two participants.

Third, as mentioned in 4.5.2.1, morphs that have different meanings that are semantically related were annotated as UCI. For example, the phonologically conditioned prefix *ḫa-* in Sumerian expresses ‘assertions, wishes, commands’. All these functions are deontic, and the sub-specifications are plausible interpretations given the grammatical context.

(12) Semantic readings of *ha/hé-* in Sumerian

a. šu **ḫé**-eb-bar-re

hand **MOD**=VP-3NHUM.on-open-3sg.A.IMPF

‘He should release it!’ (Jagersma 2010: 561)

b. ékišnuĝál-šè **ḫa**-ba-an-ku$_4$-re-en

Ekishnugal=TERM **MOD**=VP-MM-in-enter-1sg.S/DO

‘I truly entered into the Ekishnugal!’ (Jagersma 2010: 563)

In (12b), the grammatical context is a past event. Here, *ha-* expresses a strong assertion of a past action or state. The impossibility of the reading ‘wish’ or ‘command’ results from the past context. The three readings ‘wish’, ‘command’ and ‘strong assertion’ are related to the speakers’ emotional (deontic) attitude to the event and therefore semantic variants of one morph.

Fourth, morphs that co-appear with other morphs but do not change their meaning were annotated as UCI. In Mian (Ok, Papua-New Guinea), the suffix *–bio* (general past) has the following distribution:

i. "in directly inflected verbs, *-bio* can only occur in the perfective;
ii. "in final verbs, *-bio* always co-occurs with the realis marker *–n*/–Ø" (Fedden 2011: 296).

Neither the meaning nor the form of *–bio* is conditioned by perfective or realis features, which means that it was annotated as UCI.

Fifth – in addition to different, but related readings of one morph, and morphs which co-occur with other morphs but without a unidirectional conditioning relation, so-called 'free' semantic or formal variants have been also classified as UCI. These variants have no hard conditioning as they are sometime present, sometimes absent in constructions. The lack of hard conditioning is expressed in grammars differently, such as in the following passage from the Grammar of Lumun (Talodi-Heiban, Sudan):

> "[In Lumun,] [s]ingular subjects, when immediately preceding a verb or predicative adjective, are far more commonly expressed by a bound pronominal form. Use of the free pronoun, as in the examples below, is nevertheless possible." (Smits 2017: 200)

The reason for annotating morphs with free variation/weak constraints as UCI is that there is no evidence that free variation might pose difficulties in processing; in usage-based terms, free variation reflects pre-emergent exemplars that are found in use but have not yet been integrated into a system (Ellis 1999).

Sixth, morphs were annotated as UCI when there was no evidence for lexical or grammatical conditioning. Authors almost never emphasize that morphs are predictable or unconditioned. Some exceptions include examples like "In the Old Sumerian period, the modal proclitic { ḫa- } has two different forms and spellings. Their use follows clear [phonological] rules and is completely predictable." (Jagersma 2010: 558). In these cases, 'no conditioning found' was written for the positions in the spreadsheet. Negative evidence does not imply that there could be conditioning in the language which the author has not accounted for. The respective length of grammatical descriptions and cross-referentiality of topics means that some information might have been missed that is not mentioned in the main discussion of the inflectional morphs in grammars.

4.5.3.2 Distribution

UCI positions make up the minority of inflectional morphs in LCI languages (61 positions), but, together with the two control samples (GCI-sample: 28 positions; UCI-sample: 25 positions), there is total number of 114 positions. For 39 positions, there was clear information on phonological conditioning of at least one allomorph. Unlike LCI and GCI, UCI was not further classified into subtypes (such as stem alternation, grammatical categories of conditioning, etc.). The following table shows the distribution of inflectional categories across the UCI categories.

	LCI-sample	GCI-sample	UCI-sample	Total
Aspect	11	10	10	31
Tense	12	11	5	28
Mood	16	7	10	33
Indexation	25	10	2	37
Polarity	12	3	4	19
Total	76	41	31	148

Table 4.18: Distribution of categories across UCI positions (Total: 114) includes multiple functions for one position)

Table 4.18 shows that the grammatical functions are sufficiently represented across UCI positions and all samples. The most common category expressed by UCI positions/morphs in the LCI sample is indexation (32.9 %). In GCI-languages, all functions except polarity are distributed similarly among UCI-positions. In the UCI sample, the most common categories are aspect and mood (32.3 % each). This continuum between LCI, GCI and UCI languages might reflect a cline in the time depth of morphs. LCI languages are generally more synthetic and comprise different types of morphs and conditioning. Usually, lexical conditioning falls on categories that are more relevant to the stem (such as aspect), and UCI is reserved to the more recent, peripheral layers of the templates (i.e., indexation). In UCI languages, there is less synthesis, so the more common categories that are attached to verbs are the most relevant ones (aspect), while indexation is rare. Instead, grammatical relations are usually not expressed by indexation affixes, but by pronouns or noun phrases. A specific challenge when gathering languages for the UCI sample was to find enough positions with indexation without it being grammatically or lexically conditioned. In North America, which comprises numerous languages with exuberant morphology, Northern Pomo is an exception, by being synthetic and exhibiting only UCI.

4.5.4 Conclusion

The three samples provide sufficient data to compare the behavior of all categories (LCI, GCI, UCI) in relation to the stabilizing variables defined in Section 4.6. The morph positions express a variety of grammatical functions and are distributed across all positions in words (prefixes and suffixes, inner and outer positions). LCI can be part of stems; GCI morphs show a large variability in which categories they express and by which categories they are conditioned. UCI morphs behave differently across samples: in LCI languages, UCI morphs occupy the periphery of words, namely indexation and polarity. For the UCI sample, UCI expresses mostly aspectual and modal functions. This difference is one reason for why languages without LCI or GCI are necessary in deriving conclusions about the alignment of morphological complexity and other structural properties.

4.6 Stabilizing variables

Chapter III discussed functional-adaptive properties of language structure that facilitate communication, prominence and high frequency. While the psycholinguistic effect of these properties might be beneficial for memorization and acquisition by speakers, it has an inconsistent effect on the preservation of different structural elements. As argued in Section 4.3, only certain types of prominence and frequency falling on certain structural units lead to stability of structure, and these structural units might favor the stability of morphological complexity and affect its distribution. In addition, it is important to select those structural properties and units that can be easily operationalized and mapped onto the morphological structure of verbs. There is scarce data for most of the languages in the samples on the psycholinguistic effects and the discursive reality of language structure. On the other hand, insights from experimental and corpus-linguistic studies from more commonly spoken languages have been helpful in determining general behaviors of language structure. As discussed in Chapter III, there is no evidence that prominence *does not* facilitate the acquisition of language material. However, not every prominent unit is stable because it is prominent. For example, stress can change the quality of vowels (by hyperarticulation or diphthongization) or change the shape of morphs. Stress tends however to preserve the existence of the nucleus and therefore of syllables. This tendency to unaltered perpetuation of language material also pertains to high type frequency. High token frequency of units (phonemes, morphemes, words) can lead to erosion, whereas high frequency of abstract phonological or morphological schemas (i.e., high type frequency of units) typically reinforces these schemas and prevents their loss.

4.6.1 Prominence

As concluded in Chapter III, the most promising variable for studying prominence contrasts is stressed and unstressed syllables, which can be easily mapped onto morphological structure. However, this variable is not applicable to languages without stress or when morphemes are consonantal. Alternative prominent units that can be used as variables include low vowels, phonological boundaries, phonemic contrasts (such as tonal contrasts), and concrete meanings. Low vowels are useful for comparing vocalic morphemes, while consonants at phonological boundaries provide stability but require specific boundary positions. Tonal contrasts, although not directly preserving morphology, indicate stable environments when there are a large number of contrasts, making them a suitable variable for stressless languages. Lexical morphemes are more stable than affixes but their prominence due to semantic concreteness is not the cause of their stability.

This preliminary picture of prominence effects have the following implication for methodology. First, stress should be prioritized as a variable for stability due to prominence. Since stress is syllabic, and syllables always have nuclei, vocalic prominence will be ignored as a second variable for nucleus prominence. However, the close connection between nuclear and syllabic prominence means that the syllabic variable might fail to capture prominence contrasts between consonantal morphs. As determined in Chapter III Section 4.3.1.1, there is no promising connection between segmental prominence and stability of consonants, and therefore, different consonantal features might not be used to construct a stabilizing variable for consonantal morphs. In this case, it might be more plausible to use prominent boundaries as a variable to account for stability of consonants. Tonal contrasts could be employed for stressless tonal languages, even if the connection between tonal prominence and morphological stability is not as clear as with stressed syllables. Semantic prominence should be ignored since the connection between concreteness and stability is not straightforward. Thus, *prosodic* prominence is the focus in this dissertation (cf. Zingler 2020 for choosing a similar focus of examination). In the following sections, the types of prominence relevant for the annotation are discussed, and their annotation is explained in Section 4.6.1.5.

4.6.1.1 Stress

Stress is a very frequent property in the languages of the world. As such, many of the languages in the sample show stress as a prosodic property that expresses prominence. The 41 languages sample contain 33 languages where stress is mentioned as a feature in the prosodic system. Perceptually, stress usually manifests itself in a higher F0, longer vowel duration and increased loudness, or some combination of these (Lehiste 1970). From the articulatory viewpoint, stressed syllables involve more exaggerated articulation of jaw movements in relation to unstressed syllables (De Jong 1995). In most grammatical descriptions, the specific phonetic features of stress are not described in detail. Despite this, authors seem to have a good intuition of whether a language has stress or not. In the languages that did not exhibit stress, authors still make claims about which features are more prominent to speakers, sometimes with no evidence, although fortunately, there are some exceptions, such as the treatise on Navajo prominence in McDonough (2003).

The languages in the samples exhibit a variety of stress patterns. Stress can be phonologically determined (in relation to word edges, syllable count/structure, phonemes) or morphologically (specific morphemes attract stress). Table 4.19 shows the distribution of stress types across the languages of the sample.

Level of determination in verbs	Numbers of languages
Phonology alone	14 [42.4 %]
Phonology and morphology	13 [39.4 %]
Morphology alone	6 [18.2 %]
Total	33

Table 4.19: Determination of stress in languages. Morphological rules includes rules that are lexically conditioned or conditioned by specific morphemes in words.

Table 4.19 shows that most of the stress patterns are phonologically determined. In almost all languages, stress is predictable from either phonological or morphological rules. Unpredictable, lexically-conditioned stress for verbs (like in Indo-European languages such as English, Spanish or Russian) exists in two languages of the sample (Aguaruna, Mian). Because of the high degree of synthesis across the languages in the LCI sample, certain affixes and clitics can attract stress, yielding words or phrases with multiple stressed syllables. In Nunggubuyu, long vowels and penultimate syllables attract stress, which in some cases leads to having two stressed syllables in one word, such as *ná:náni* 'We saw them' (Heath 1984: 32).

Because stress is widely distributed across the languages of the sample, this property was prioritized in the annotation of prosodic prominence. In annotating stress, one relies on the assessment of what authors consider to be prosodically prominent, without having to analyze how specific phonetic features interact (high F0, loudness, specific phonemes). Languages that have lexical stress may have more than one stressed syllable, but in most cases, there is a single syllable that is considered the most prominent, referred to as "primary" stress (Hayes 1995). This means that a single most prominent syllable can be identified, making stress an effective variable for showing alignment with the morphological template. If stress was distributed equally across multiple syllables in words, the alignment between (potentially) stressed and unstressed syllables and morph positions would lower the significance of the results.

When stress is phonologically determined, it might fall on several morphological positions of the template. In such languages, measuring the relative likelihood that specific morphs align with the prosodic structure is dependent on the relative frequency of alignment, which most grammars do not provide. Without enough data on the language, the alignment between phonological and morphological structure must be deduced by its rules and the patterns of obligatoriness of morphological positions.

For example, in Azari (Turkic; Iran), stress usually falls on the last syllable of the word (Lee 1996: 17). This means that any suffix after the obligatory morphs has the potential to be stressed, and the last suffixes of the verb are always stressed. However, one must also consider the phonological shape of suffixes. With regards to Azari, the Appendix shows that the morphological structure has been divided into two stem positions and 10 suffix positions. Positions 6-10 have been annotated as potentially attracting stress ('possible'), whereas positions 1-5 have been annotated as never attracting stress ('no'). Position 10 has been annotated as obligatory, which means that if this position contained only syllabic morphs, stress would only fall on morphs of this position. The indexation markers of Position 10 are largely monoconsonantal, (such as *–m*; 1sg) but the suffix *–lər* '3pl', for example, is syllabic. In this case, only *–lər* is always stressed (*gə.lə.cək.'lər* 'they will come,' Lee 1996: 17), whereas in the case of *–m* (*gəl.'dim* 'I came,' Lee 1996: 47), stress falls on the vowel of the preceding morph (*–di*: past tense) from Position 9. Morphs not being syllabic in Position 10 is the reason why stress can fall on preceding morphs, and these morphs occur from positions 6 through 9. Morphs from positions 1 through 5 are never part of the last syllable of words. In addition, morphological rules for stress placement exist also in Azari. For example, the second person imperative *–(y)In* from Position 10 never has stress, even if it is in final position (Lee 1996: 17).

Thus, to deduce the morph-stress alignment one must consider the stress rules and generalize them if not further specified. Further specifications might exist due to lexemes or morphemes attracting stress. Secondly, one has to consider the shape, position and obligatoriness of the morph, as well as the potential for surrounding morphs to change the environment for stress. The specific procedure in annotating this morpho-phonological alignment is further discussed and exemplified in Section 4.6.1.5.

4.6.1.2 Tone

15 languages in the sample have tonal contrasts. The minimum number of contrasts is 3 (Navajo, Skou, Koasati, Kiowa, Nyamwezi, Bangime), the maximum number is 9 (Luo). The number of contrasts for each language is listed in Table 4.20.

Language	Number of tonal contrasts	Reference
Bangime	3	Heath & Hantgan (2019: 29)
Ket	4	Vajda (2004: 8)
Kiowa	3	Watkins (1980: 34)
Koasati	3	Kimball (1991: 26)
Lumun	4	Smits (2017: 81)
Luo	9	Tucker and Creider (1994: 43)
Mian	5	Fedden (2011: 50)
Navajo	3	McDonough (2003: 6)
Nyamwezi	3	Maganga & Schadeberg (1992: 40)
Pirahã	4	Everett (1986: 312)
Sheko	4	Hellenthal (2010: 111)
Skou	3	Donohue (2004: 54)
Supyire	4	Carlsson (1994: 42)
Zacatepec Chatino	5	Villard (2015: 148)
ǂHòã	5	Collins & Gruber (2014: 11)

Table 4.20: Number of contrasts in tone languages of the sample.

The number of contrasts in the languages listed in Table 4.20 is determined on the basis of phonetic contrasts, which might be different from the phonemic/tonemic inventory. Some authors propose fewer underlying contrasts, such as Everett (1986). In other cases, the number of phonemic tones derives from combination of other tones (such as in Navajo or Kiowa). Table 4.20 shows that there is a considerable number of tonal languages in the sample, especially in the LCI sample (12 out of 31, or 38.7 %). This reflects quite closely the proportion of tonal languages in the world, which is 41.7 % in the global sample of Maddieson (2013b), which consists of 527 languages.

In Chapter III (Section 3.3.1.4), it has been argued that a higher tonal variability in a position and more marked tones are more prominent to speakers than frequently occurring tones. Furthermore, tones that differ the most in their F0 height tend to be retained (such as the lowest and highest tone), unlike tones that have a similar frequency in contrast to others. The prominence established through tone can be syntagmatic (a tonal syllable in a word catches attention in contrast to neighboring syllables) or paradigmatic (a tonal syllable catches attention for being novel/occurring infrequently in the discourse). Thus, one could say that prominence of tone differs from prominence of (predictable) stress by an additional paradigmatic dimension, although this still needs to be investigated.

However, it has been noted that tonal prominence might only be a cause for tonal stability, not morphological stability. Instead, it is more likely that another environment is responsible for the stability of both the tonal contrasts and morphological complexity, such as word stems. Nevertheless, given that there are no studies that convincingly show the causation relationship between tonal prominence and complexity, I decided that it is still worth analyzing whether tonal prominence is associated with specific conditioning of inflectional morphs. After all, tonal prominence might be related to syllabic prominence/stress, and Goldsmith (1987) has shown that the evolution of tone might be connected to the evolution of stress.

The challenge for annotation consists in determining how prominence through tone is aligned with the positions of verbal templates in the languages of the sample. Tonal contrast is a relational category, and morphs can only have one tone in each context. The best case to establish high tonal variability for one morph is if the morph shows a great array of tones in many phonological conditions. However, in many languages, certain morphs use only a subset of tonal contrasts, such as in Navajo or Zacatepec Chatino, while the full variety of contrasts is present only in stems.

One solution for this problem is to consider those morphs with tones that have the lowest or highest relative F0 value. These morphs could be considered more prosodically prominent than morphs with tones that differ less in F0 height. A second solution would be to annotate those morphs as prominent with tones that are not frequent, therefore more marked. This would imply knowledge of the frequency of occurrence of specific tones, and this frequency can be rarely obtained from grammars. A third solution is to annotate tone variability across morphological paradigms, i.e., the number of possible tones that the morphs in a paradigm carry. This solution suits the present study, since prominence values are only established based on paradigmatic positions, not single morphs. Tone variability in paradigms suggests that there is tonal prominence in relation to morphological schemas; this means, paradigms are salient to speakers because the morphs of these paradigms have diverse tones that vary from construction to construction. This relation between tonal contrasts and morphological paradigms might sound plausible but still needs to be psycholinguistically investigated. There might be other advantages in annotating tone variability of paradigms: if the variability of tones is high, at least one morph in the paradigm contains the tone that is most prominent. On the other hand, if all the morphs in a position have the same tone, it is less likely that a specific morph would be more prominent than another morph from the same position; however, they may still be syntagmatically prominent. For example, if all the morphs in one position have a high tone, they would prosodically contrast with adjacent morphs that have a low tone. Thus, while rather suggestive, variability of tone among morphs in one position might be one way to capture both prominence derived from tonal contrast and prominence derived from F0 height; however, this does not apply to cases where F0 height is dependent on discourse contexts such as different intonations in statements and questions.

Another challenge for annotation is how to operationalize high tonal variability among languages with different numbers of tones. In contrast to stress, where a tripartite distinction between morphs and stress can be applied to all languages (morph stressed, morph sometimes stressed, morph unstressed), tone prominence entails different gradient distinctions relative to the number of tone contrasts. For example, a morphological paradigm that has only two tonal contrasts is less prominent than a paradigm that has three tonal contrasts, and this paradigm might again be less prominent than a paradigm with five tonal contrasts. In a language with only three tonal contrasts, there is less gradience than in a language with more tonal contrasts. This language-specific gradience is not easily operationalizable for all languages. Therefore, selecting and generalizing prosodic prominence based on tone is more difficult than based on stress.

Despite these issues, tonal contrasts could be chosen as an alternative annotation when a language does not have stress or clearly-defined prosodic prominence. In the languages that have both tone and stress, it is easier to argue that stressed syllables are more prominent than a specific tone/tonal contrast. Of the 15 languages that have tone, 8 do not exhibit stress (Ket, Zacatepec Chatino, Sheko, Koasati, Kiowa, Lumun, Nyamwezi, ǂHöã). As such, values indicating number of contrasts have been annotated for these languages. These values are further explained in 4.6.1.5.

4.6.1.3 Phonotactic position

Further prominent properties discussed in Chapter III include phonological boundaries, such as initial or final phonemes in syllables or words. For convenience, the cline between most and least prominent phonotactic positions excerpted from Ségeral and Scheer (2008) and presented in Section 3.3.1.3 is shown once again:

strongest position *weakest position*

word-internal onsets after obstruent codas >
word-initial onsets >
word-final codas >
word-internal codas >
word-internal intervocalic onsets.

In total, 24 inflectional positions in the sample contain morphs that are purely consonantal (20 positions in the LCI sample, 3 in the GCI sample, 1 in the UCI sample), which means that stress or tonal contrasts cannot be directly applied as criteria for determining whether these morphs are prominent. For these consonantal morphs, phonological boundaries might be a better predictor for prominence and a variable along the cline shown above could be constructed.

However, this cline shows that operationalizing positional prominence in order to align it with the morphological template might be quite a difficult enterprise. Determining the alignment between morphs and 'word-internal onsets after obstruent coda', 'word-internal codas' and 'word-internal intervocalic onsets' would require a careful investigation of the possible segments of morphs and their likelihood to occur next to one another across verbs. This task has been regarded as too tedious as it would involve splitting up the morphological positions into morphs and segments or annotating the phonological features of each consonantal morph.

On the other hand, the alignment between consonantal morphs and word boundaries ('word-initial onsets' and 'word-final codas') is easier to determine. The general potential for one morph to appear at a word boundary depends on whether other morphs precede or follow it. If no morphological position is obligatory (except the stem), then any morph might appear at the end or beginning of a word. Consonants at word boundaries would be generally more prominent than consonants in the middle of words, except for word-internal consonants after obstruent onsets. However, even by ignoring this latter position the procedure to derive the likelihood for a consonantal morph to appear in a prominent position misses the factor of frequency. For example, it might be the case that a morph from an outer prefix position might not occur as frequently than a morph from an inner position. This is the case in Navajo. The verbal template in Navajo exhibits 10 prefix positions (as analyzed in this dissertation). Verbs do not always have prefixes from the outer positions (-10, -9, -8) but they always have prefixes from the Position -2. Consequently, Position -2 prefixes might overall appear more often at word edges than prefixes of positions -10, -9 or -8.

In addition, as was argued in Chapter III, word boundaries do not accurately predict the retention of morphs over time; while word-initial onsets are not only prominent but also stable, word-final codas are not necessarily more stable. However, intervocalic consonants are the least stable. Nevertheless, as Echols and Newport (2012) note, word-final and stressed syllables are omitted by children at a lower rate than non-word-final and unstressed syllables, and this evidence suggests that word-final consonants, when they are part of stressed syllables, could be also more stable than word-medial consonants that aren't part of stressed syllables. As further detailed in Section 4.6.1.5, a combination of stress and position was chosen in order to account for prominence of consonantal morphs.

4.6.1.4 Non-prosodic prominence

As mentioned at the beginning of Section 4.6.1., I decided not to annotate morphological, semantic or pragmatic prominence of morphs. Chapter III did not provide enough evidence for meaning-based prominence playing a role in the preservation of structure. Prominent morphs, such as stems, can turn into grammatical morphemes, and lose their prominence. Likewise, semantic prominent features might not resist change either since they can easily become more general and lose specificity. Pragmatic prominence such as focus is not bound to a specific form; instead, it can fall on different constituents.

Furthermore, there would be challenges in obtaining significant results when aligning morphological structure to meaning-based prominence. With regards to the fact that stems are semantically more prominent than affixes due to their concreteness, the results would be straightforward: lexical inflection is more likely to be semantically prominent since it is almost the only conditioning type that is expressed by stems. Investigating the effect of semantic prominence would again yield a trivial distinction between stems and affixes. With regards to pragmatic prominence, verbs are less likely to be part of focus constructions, and if they are focused ('predicate focus'), there are several constructions, reflecting their marked character (Zimmermann 2016); however, pragmatic focus does not distinguish between smaller parts of words (i.e., affixes are almost never focused by speakers). As such, focus does not lend itself to contrasting pragmatic prominence between morphs. In conclusion, the best solution is to contrast the morphological structure of a verb with prosodic prominence.

4.6.1.5 Annotation and distribution of prominent positions

In order to identify prominent positions that could account for structural stability, it was desirable to use an annotation that reflects different degrees of prosodic prominence. Because annotating stress is straightforward, it was prioritized for annotation purposes. In fact, stress is always prosodically prominent, whereas a specific tone is not necessarily. Thus, positions of morphs in stress languages were annotated according to whether they exhibit morphs that are stressed or not. A tripartite distinction between 'yes', 'no', and 'possible' was chosen to account for positions with morphs that are always stressed, positions that are never stressed, and positions whose morphs are stressed in certain constructions. The value 'possible' was chosen for when the entire position shows inconsistent alignment with stressed syllables or when the morphs of the position exhibit different values (such as morph 1 is always stressed, morph 2 sometimes, morph 3 never). While quantifying the proportion of the morph-stress-alignment in a position is feasible, time restrictions did not allow for conducting this analysis. Following the procedure outlined in Section 4.4.2.3, if one position were to contain morphs that show a different stress placement *and* belonged to separate categories/paradigms, the position was split up and independent values were assigned. The assignment of these values is exemplified in the following excerpt from the annotation of Pilagá (Guaicuruan, Argentina) (Vidal 2001).

Position	Morphs Allomorphs are listed separated by slash marks.	Promi-nence	Explanation	Examples (stress is marked with acute accent)
-2	*qo–* (indefinite subject) (p. 146)	no	This prefix is never stressed. (p. 71)	***Qo***-*n-yáʕana* **IS**-setB.3-call 'They called him.' (p. 70)
2	*–(a)q*/ *–soq*/ *–ʕa*/*–sa* (1pl) *–i*/*–e*/*–q(a)e* (2pl) *–di*/*–y* (3pl) (p. 150, 152)	possible	Some of the subject plural markers (2pl) can be stressed, others not (p. 160–161)	*an-qač-**í**-ñi*[16] setB.2-catch-**2.pl**-DIR.downwards 'You all caught.' (p. 161). *na-qáč-i-**di**-ñi* set.B.3-catch-EPV-**3pl**-DIR.down-wards 'They caught.' (p. 161)
6	*–a* (object singular) *–to* (object paucal) *–lo* (object plural) (p. 165)	yes	Object number suf-fixes always carry stress (p. 72–73)	*n-ibiet-'at-a-**ló***[17] setB.3-control-RECP-ep.vowel-**OBJ.pl** 'They control each other.' (p. 165)

Table 4.21: Annotation of prominence values for selected positions of the Pilagá verbal tem-plate.

Table 4.21 shows examples for these three values: In Position 6, every morph in the position can carry stress (suffix position 6); In P-2, no prefix carries stress; in P2, some morphs carry

[16] In Vidal (2001: 161) *–í* is analyzed as an epenthetic vowel, such as in *na-qáč-i-di-ñi*, but the author notes that "there is a stress shift in the plural verb form." (p. 160), indicating that –*í* belongs to the indexation morphs listed in the second suffix position in Table 4.21.
[17] Stress is not indicated in the example, but is explicitly stated for *–ló* in Vidal (2001: 73)

stress and others do not (P2). In Pilagá, stress is lexically assigned, with several phonological and morphological rules (See Vidal 2001: 70–74). While only some of the indexation morphs attract stress in P2, this position was not split because all these morphs belong to the same paradigm (subject plural). As such, lexicalized, fixed, variable and multiple conditioned stress patterns were subsumed under a uniform tripartite distinction of prominence.

What happens to the languages with so-called 'weak stress' or no stress at all? Some grammatical descriptions from the sample stated either that stress was not relevant, or they did not mention stress at all, such as in Lumun (Smits 2017) or Bangime (Heath & Hantgan 2018; Hantgan 2013). In other grammars, prominence is associated with a specific tone. However, all languages whose grammars do not mention stress are tonal. As such, tonal contrast could be exploited for the annotation. As argued in 4.6.1.2, a higher number of contrasts in one position suggests a higher probability that one morph in the position is prominent.

To align prosodic contrasts through tone with stress, zero contrasts in one position were annotated as 'no', maximal contrasts as 'yes,' and some but not all contrasts as 'possible'. This has been done for most tonal languages without stress (Sheko, Zacatepec Chatino, ǂHȍã). For the languages with only three tonal contrasts, I decided to rely on the authors' knowledge of which tones are more prominent or can be compared to stress. If this information was not available, I have chosen the highest and most marked tone as the most prominent one; that is, if words appeared to exhibit fewer high tones than neutral or low tones, I chose the highest tone. This has been done for Kiowa, Koasati, Lumun and Nyamwezi. In Bangime, low and high tones were about equally distributed (a word could have a succession of high and low tones). In this case, a specific tone would less likely act as a prominent property that stands out from another. Here, I decided to choose a feature associated with stress, namely mora length:

long moras in Bangime are less frequent than short moras. Consequently, long moras were considered prominent, and short moras as non-prominent. The following flowchart summarizes the process of annotating the three values ‘yes’ ‘no’ ‘possible’ for morphological positions with syllabic morphs.

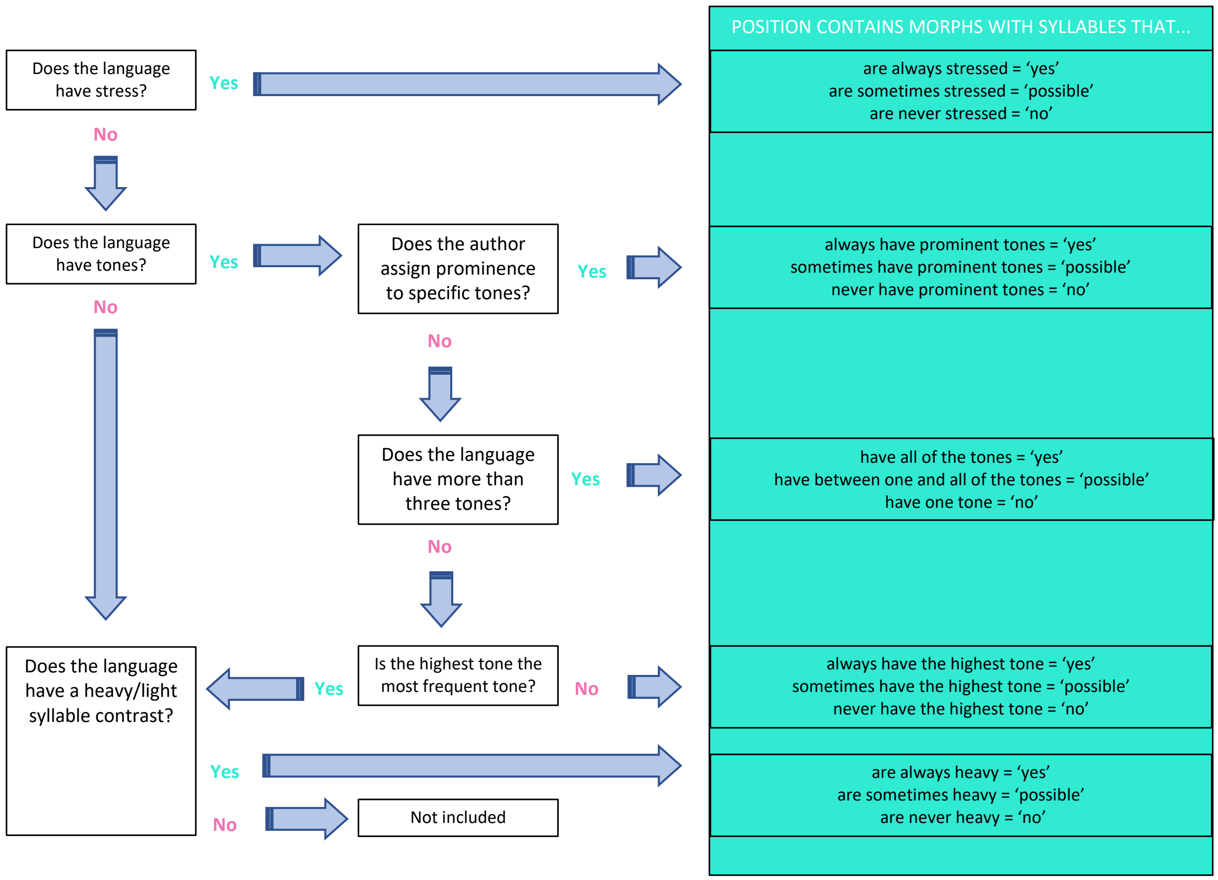


Figure 4.4: Selection process for assigning prominence values to stress, tone and weight-sensitive languages. ‘yes’ = always prominent; ‘no’ = never prominent; ‘possible’ = sometimes prominent. This flowchart has been used for syllabic morphs.

The selection process for positions with syllabic morphs can be exemplified in the following languages.

Dumi Position 0 (stem) (Van Driem 1993)

- Dumi has stress (p. 58).
- "Verbs (...) are always stressed on the stem" (p. 58).
- Annotate Position 0 as 'yes.'

Sheko Position 3 (subject) Morphs: *–n* (1sg, syllabic) *–ha* (2sg) *–há* (3sg.M) *–yí* (3sg.F) *–ń* (1pl, syllabic) *–ít(i)*/*–íʃi* (3pl) (Hellenthal 2010: 429)

- Hellenthal (2010) does not mention stress in the grammar.
- Sheko has tone (p. 111).
- Author does not assign prominence to a specific tone
- Sheko has more than three contrasts (1, 2, 3, 4) (p. 111)
- Morphs in position 3 exhibit up to three contrasts (v́, v̀, v, but not v̄)
- Annotate Position 3 as 'possible.'

Kiowa Position 1 (aspect). Morphs: *–Ø/–ɔ́/–iá(y)/–é/* (perfective) (Watkins 1980: 206).

- There is no explicit stress that is independent of tone in Kiowa
- Kiowa is tonal.
- Watkins (1980: 49) mentions that 'syllables with high and falling tone are more stressed than those with low tone, a long syllable is more stressed than short one, length combined with high or falling tone makes stress particularly marked.'
- Tones are prioritized in annotating prominence. Morphs in Position 1 all have a high tone.
- Annotate position 1 as 'yes'.

Bangime Position -3 (aspect, mood, polarity). Morphs: *daw*/*dáà*/*dá*/*nà*/*ndà* (incompletive) *kama*/*kóò* (completive) *hà*/*há* (irrealis) (Hantgan 2013: 252) *bié* (negative) (p. 321) *maa* (prohibitive) (p. 238) Ø (perfective)

- There is no mention of stress patterns in the grammar.
- Bangime is a tonal language (p. 70).
- There is no mention about which tone is more prominent.
- Bangime has 2 or 3 tones: high and low according to Hantgan (2013: 70), and high, mid and low according to Heath & Hantgan (2018: 29).
- There is no information whether high, mid or low tone is the most frequent one; in fact, they seem to be equally distributed across syllables.
- Bangime makes a distinction between long and short moras (Heath and Hantgan 2018: 28), which can be interpreted as a contrast in heavy and light syllables.
- Some morphs in Position 3 have long moras (*dáà*, *kóò*) others are short (*hà*, *há*).
- Annotate Position –3 as 'possible'.

With regards to positions that only exhibit consonantal morphs, I decided to include two prominence properties to ensure that the consonantal positions annotated as 'yes' exhibited a reliably high degree of prominence. The 'yes'-value was assigned to all the consonant morphs that could appear in initial or final positions of words *and* were either onset or coda of prominent syllables as defined in Figure 4.4. Consonantal positions that were not part of prominent syllables were annotated as 'no' for prominence. Consonantal positions that never appeared as the initial or final position of words were also annotated as 'no' in the spreadsheet. If consonantal positions always appeared as parts of prominent syllables, they were annotated as 'yes'

– if they at least had the possibility to be the initial or final consonant. Consonantal positions were annotated as ‘possible’ if they sometimes were part of prominent syllables, and sometimes or always initial or final consonants. Consonantal positions were annotated as ‘no’ if they were never part of prominent syllables and/or never appear in the initial or final position of words. This choice is based on evidence that the interaction of stress and position is a very salient property to speakers (Echols and Newport 1992). The flowchart for the selection of annotation values for consonantal positions is presented in Figure 4.5.

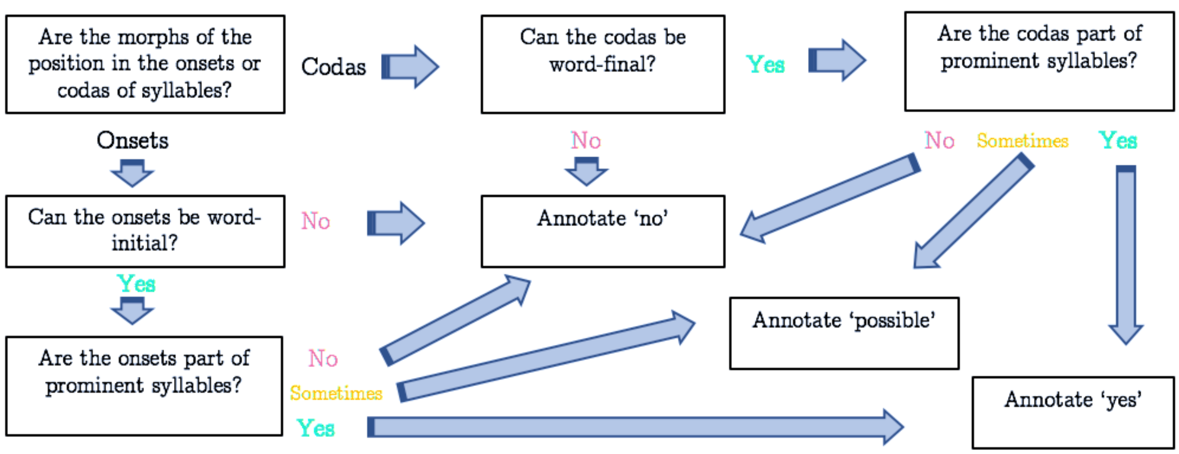


Figure 4.5: Selection process for the annotation of positions with purely consonantal morphs.

The selection process for positions with purely consonantal morphs can be exemplified in the following languages.

Ingush Position 0 (gender). Morphs: *j*– (J), *v*– (V), *b*– (B) *d*– (D) C– (stem consonant if there are no gender morphs).

- Morphs appear as onsets of syllables.

- Morphs can be word-initial, such as in ***d****woaxdu* 'heat up' (Nichols 2011b: 301)
- Morphs can be stressed since stress is marked almost exclusively on the first syllable (p. 96) /'dwoax.du/ 'heat up.'
- Annotate position as 'possible.'

Tunisian Zuwara Berber Position 5 (polarity). Morphs: *–š*/*–Ø* (negative)

- Morph *–š* appears only in codas (zero morph ignored)
- Morph can occur in word-final position, such as *wəssinə́ɣš* 'I don't' (p. 104)
- Morph is always stressed since "negative suffixation attracts the accent and all negative forms are oxytones" (p. 91). Since this is the last morph of the template, and negative words are always stressed, this morph is always part of the coda of stressed syllables.
- Annotate position as 'yes.'

It should be noted that positions that contain both consonantal and syllabic morphs were encoded according to the behavior of syllabic morphs. That is, consonantal morphs were ignored when a position contained other syllabic morphs. As such, the annotation of syllabic morphs was prioritized.

Table 4.22 shows the distribution of all positions containing syllabic morphs (including non-inflectional positions).

Prominence Value	Yes	No	Possible	Total
Number of syllabic positions	55	189	196	440
Number of positions with purely consonantal morphs	5	11	18	34
Total	60	200	214	474

Table 4.22: Count of prominence values across all 474 positions of the sample (inflectional and non-inflectional[18]) containing syllabic morphs in all three samples.

As Table 4.22 shows, the distinction between 'yes' (always aligned) 'no' (never aligned) and 'possible' (sometimes aligned) is appropriate in grasping the diversity of position-prominence alignment. Since most of the syllables are unstressed, it is understandable that 'yes' values are the minority. Despite this discrepancy, the distribution provides sufficient data to investigate the correlations with the different types of complexity. The results of these correlations are presented in Chapter V.

4.6.2 Obligatoriness

The second structural property that was argued to be responsible for a longer lifespan of language structure is strong entrenchment, which is dependent on high frequency. However, not

[18] If the position contained inflectional morphs that are purely consonantal, but also contained non-inflectional morphs that can be syllabic, the position was still annotated as purely consonantal, since only inflectional morphs are targeted.

every highly frequent unit is preserved. In Chapter III, it has been concluded that frequency of both phonological and morphological schemas (high type frequency of patterns) has a self-preserving character which is also responsible for the preservation of structure that aligns with these schemas. Morphological schemas result from the frequent co-occurrence of a set of morphs with another set of morphs. For example, the morphological schema STEM+TENSE in English results from present and past tense markers always being in a position that follows the stem. High type frequency of the tense morph reinforces the schema. Phonological schemas emerge in a similar fashion: the frequent co-occurrence of sets of phonemes in a specific order (such as the succession of vowels and consonants) reinforces the schema and leads to the accommodation of phonemes that do not fall into the schema. Examples for the conservative character of frequent phonological schemas are given by Blevins with regards to Oceanic languages (Blevins 2009). In contrast, high frequency of words, morphemes and phonemes (that is, high token frequency of elements) can lead to erosion. The question for this methodological chapter is how high frequency of morphological and phonological structure can be operationalized.

In 4.4.2., it was proposed that obligatoriness of morphological categories and syllables reflects a higher type frequency compared to other morphemes and syllables. But how can obligatoriness be determined as a variable for this study? In the following, I will explain what type of obligatory units lend themselves to annotation.

4.6.2.1 Morphological obligatoriness

Morphological obligatoriness can be understood in two ways. One way of thinking about obligatory morphemes is that specific grammatical contexts always require the marking of a category. The grammatical contexts can be broad or narrow. For example, one can say that in German, every sentence has an overt subject, a noun phrase or a pronoun. The grammatical context is broad, because ‘sentence’ is a frequent unit in discourse. Another example is: “in English, aspiration is obligatory for voiceless stops before vowels.” The context here is quite narrow, namely stops at the beginning of words. Obligatoriness can therefore relate to any predictable rule in a language system given a specific context.

Since in this study, obligatoriness should reflect high frequency, the contexts that are searched must be as broad as possible: contexts which occur very frequently in speech. From a usage-based perspective, obligatoriness is an emergent phenomenon of high type frequency, that is, the occurrence of an entrenched schema as part of another deeply entrenched schema. Obligatoriness here relates to entrenched phonological or morphological patterns of the schema ‘verbal construction’, which is itself deeply entrenched. This conception of obligatoriness is based on the contrast between more and less entrenched schemas, and therefore between more and less frequent units. Since verbs are the most frequent parts of speech in clauses, obligatory positions in verbs are among the most entrenched morphological units in the mind of speakers.

Morphological obligatoriness as understood schematically does not refer to the constant occurrence of one specific morph in verbal constructions but to the constant filling of a schematic position by morphs of a category. The targets are thus positions that are always filled in verbal templates. Whenever authors note that ‘every verb requires a tense suffix,’ the position that expresses tense in a verbal template can be characterized as obligatory. However, in most cases, paradigms also have zero-marked suffixes. From the formal side, zero morphs make a position not obligatory as they lead to a weaker entrenchment of a phonological pattern relative to a morphological pattern. However, from the semantic side, one could argue that zero-morphs do not lead to a weaker entrenchment of semantic categories since their semantics are mostly clearly defined (although not in the case where zero morphs have facultative functions; for example, when words not overtly marked for singular denote plural referents).

According to Gerner and Ling (2020: 2), zero morphs are always part of paradigms, and thus the product of highly entrenched morphological schemas. Zero morphs are the exception within a schema that presupposes the existence of formal elements. While zero morphs are stated on grounds of obligatoriness, they simultaneously suggest non-obligatoriness of a paradigm for certain features. This paradox, however, should not affect the decision of whether paradigms with zero morphs should be annotated as obligatory. From a usage-based perspective, zero morphs can reflect high morphological frequency, if zero morphs have developed from reduction processes of high frequency. Zero morphs might also reflect high type frequency of mutually exclusive morphs in the same paradigm, since zero morphs are often established in paradigms that can be applied to many words, justifying the need to mark the absence of a morph as zero. In any case, the paradigm that zero morphs belong to are very likely highly entrenched. What is more relevant is whether the category that the zero morph expresses is itself is obligatory. That is, if subject marking is present in every verb construction, one could annotate subject as obligatory. On the other hand, if only a subset of verbal constructions mark one inflectional category (such as direct object affixes), this category is *not* obligatory for all verbal constructions (it might be only obligatory for the narrower context of transitive verbs). Because the purpose of using ‘obligatoriness’ as a variable is to operationalize high frequency, obligatoriness should be defined at the broadest level possible (here: all verbal constructions). Zero morphs, if they are stated for a few gaps in the paradigm, do not decrease the entrenchment of a morphological paradigm, rather, they might be a symptom of deep paradigmatic entrenchment.

Obligatory paradigms can of course lack zero morphs. The most common obligatory position is the stem of words, but the difference between obligatory stem morphs and affixes of one paradigm is that the meaning of stem morphs is maximally different from other stem morphs, whereas affixes in the same paradigmatic position are usually semantically similar (i.e., they belong to the same grammatical category).[19] Thus, while stems are obligatory in the phonological sense (words always need a stem), their paradigmatic variability leads to weak semantic entrenchment of the stem position. However, with inflectional positions expressed through stem alternation, the obligatory phonological aspects of stems can benefit the entrenchment of these morphological alternations. Obligatory paradigms without zero morphs not only lead to entrenchment through generalization on the semantic side of morphs, but also on the phonological side.

In sum, obligatoriness values can be straightforwardly assigned to each position of the verbal templates in the spreadsheet and can be contrasted with their complexity values. Before explaining the procedure that determines these values, a look at phonological obligatoriness is needed.

[19] Of course, a clear distinction between affixes and stems is not always feasible for all languages. Languages with incorporation and a high number of lexical affixes might reduce the contrast of paradigmaticity between stems and bound morphemes. According to Croft (2001: 270), what is usually called 'stem' is the 'primary information-bearing unit' (PIBU) which is defined as "whichever form is in paradigmatic contrast with more elements".

4.6.2.2 Phonological obligatoriness

Syllabic requirement can be an alternative variable to account for frequent, entrenched patterns. Chapter III Section 4.4.1.1 has provided evidence for the effect of phonological entrenchment by showing that children insert phonological chunks that approximate morphemes but are phonologically plausible. For example, in Navajo the target word *iih yíní'ą́* 'you put it inside it.' was pronounced *hííní'ą́*, according to Chee (2017: 364). While the form *hííní-* differs from the target morphemes *iih* (inside) and *yíní-* (2sg.yi-perfective), it is identical to the form *hííní-* (2sg.seriative.ni-perfective) and is thus a phonologically conventional pattern. Peters (1997, 2001) called these phonological strings 'proto-morphemes'. While in some languages, certain 'peg-elements' (such as the 'o' in *speed**o**meter, blog**o**sphere*) might reflect entrenched phonological patterns, it is hard to utilize this evidence for the current investigation. Peg elements are analyzed as formal elements without discernible meaning that uphold a phonotactic pattern. As such, they are the opposite of zero morphs: for zero morphs, a form is presupposed, but is not expressed; for peg-elements, a meaning is presupposed, but cannot be defined.

Zero morphs reflect high frequency of morphological schemas (i.e., paradigms), whereas peg elements reflect high frequency of phonological schemas (i.e., phonotactics). However, while zero morphs are frequently found across the languages of the world, peg elements are rarer. In Navajo, the peg element *yi-* appears to fill the minimal bi-syllabic requirement of verbs if no other prefix is present. However, for a morph to be analyzed as a peg element, the contexts in which it occurs must be diverse, so that it is not associated with a specific meaning. In Navajo, the only context where this peg element occurs is when there is no syllabic prefix attached before the stem. This can happen when the morph *sh-* for first person or the zero morph *Ø-* for third person is attached, and no other prefixes. Thus, the combination of *sh-* ‘1sg’ and *cha* ‘cry’ renders *yishcha* ‘I cry,’ and the combination of *Ø-* ‘third person’ and *cha* ‘cry’ renders *yicha* ‘s/he cries.’ An alternative analysis that avoids the reification of a peg element is to analyze *yish–* as ‘1sg.imperfective’ and *yi-* as ‘3sg.imperfective’, i.e., as cumulative morphs, since imperfective can be zero-marked. This analysis is less abstract. Thus, in most cases, peg elements can be interpreted as a morph that entails grammatical information rather than a morph that is conditioned by morphophonological factors. For this dissertation, peg-elements that can be analyzed as grammatically conditioned morphs have been named ‘formatives’, and peg-elements that can be analyzed as phonological insertions were included as phonologically conditioned morphs.

Thus, while peg elements reflect high phonological entrenchment, they can almost always be analyzed as individual morphs or parts of morphs. Thus, defining and operationalizing phonological obligatoriness can be avoided. In fact, obligatory syllabic requirement can be only stated in relation to a morpheme position. Otherwise, it would not be easy to determine which part of a word pertains to the obligatory syllable, and which part is ‘added’ to the requirement. Languages like Navajo, which require an obligatory syllable besides the stem, are rare, and only a few languages in the sample can be said to exhibit rules for phonological material acting as peg-elements to fill a potential position in the template.

However, phonological obligatoriness can still be accounted for, namely when one position is always filled in contrast to a position that is obligatory but contains zero morphs. Phonological obligatoriness could be applied as a property of already obligatory morphological paradigms. When morphs appear very often in the same position (relative to other morphs), this strengthens both the phonological and morphological representation in the mind of speakers. For example, in Navajo, the bisyllabic schema containing at least one prefix and a stem is strengthened by the restriction that monosyllabic verbs containing one stem do not exist; the phonological obligatoriness of a pre-stem syllable also explains why peg-elements are present in the Navajo language. Since positions in the template are morphophonological generalizations, obligatory filling of a position with morphs can account for phonological and morphological obligatoriness. The relevance of phonological obligatoriness within morphological positions (i.e., obligatory morphological paradigms without zero morphs) is taken into consideration in the next section.

4.6.2.3 Annotation and distribution of obligatory positions

The division into morphological positions is useful in three ways: it enables the determination of paradigms and conditioning, the mapping to prosodic prominence, and also has been shown to be useful for annotating obligatoriness. All 474 positions from the three samples were annotated according to whether their morphs were present in all verbal constructions, whether their paradigm was present in all verbal constructions but included zero morphs, or whether they were not present in all verbal constructions. But what does 'all verbal constructions' mean in practice? I decided to annotate those positions as obligatory which occurred in every finite verb – that is, verbs in main clauses. The reason is that finite verbs are the most frequent type of verb. One exception has been made: I excluded zero-marked imperative constructions. In many languages, imperative constructions are the only ones that lack indexation, tense, or aspect. If imperative constructions were included in the determination of obligatoriness, almost no inflectional morphs would be counted as obligatory, and this would make the operationalization of high frequency impossible. However, I did not exclude overt imperative morphs when they occurred in a position also used for non-imperative constructions, since this is also a sign of deep entrenchment.

For a position to be associated with deep entrenchment in the mind of the speaker, it must not only be frequently filled by morphs, but the category of morphs must be consistently aligned in linear order. Morphological paradigms that are distributed across several positions will less likely lead to the association of a specific category with a specific position, and therefore, involve weaker entrenchment of a position. An example of a category distributed across several positions is the Future in Navajo (example 13):

(13) Distributed marking of Future in Navajo (Young & Morgan & Midgette 1992: 910)

dí$^{-3}$-n^{-3}**-éesh**$^{-2}$**-daał**^{0}

INC-down-**PROG.1sg-sit.FUT**

‘I will sit down.’

In Navajo, Future is expressed by three morphs from different positions: inceptive in P-3, progressive in P-2 and Future through stem alternation. Thus, for a category to be annotated as obligatory, all morphs of the category need to appear in the same position. However, a position was still annotated as obligatory if it was always filled by a morph or ‘peg element’ in the abovementioned context. In Navajo, P-2 is obligatory, as it is filled by other TAMI-morphs.

Given this relationship between peg and zero morphs discussed in Section 4.6.2.2., an annotation was chosen that would reflect deep entrenchment of a position. Thus, for positions that were always filled with morphs, the value ‘yes’ was chosen, for positions with obligatory grammatical categories but with few zero morphs the value '(yes)' was chosen, and all other cases were annotated as 'no'. More details in the selection of annotation values are presented in the flowchart in Figure 4.6:

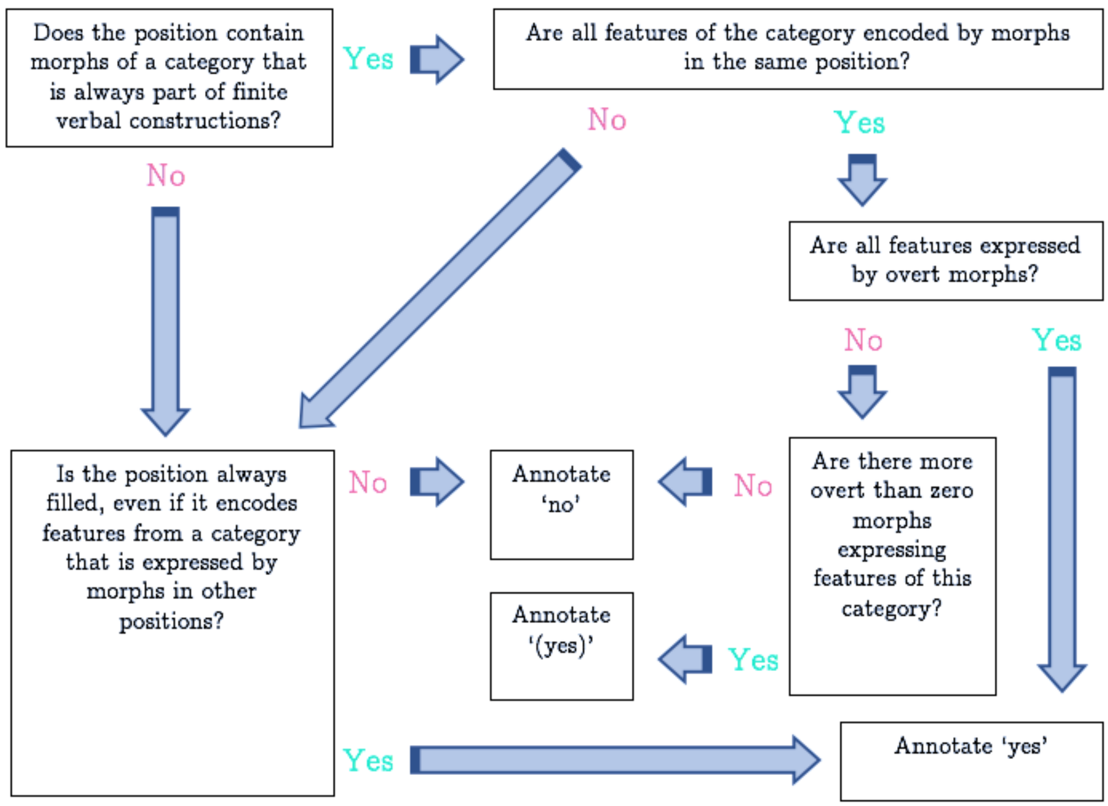


Figure 4.6: Selection process for annotating obligatory positions.

The selection process for obligatory positions is exemplified through the following positions in languages:

Mian Position 4 (subject indexation). Morphs: *–i* (1sg) *–eo/–ebo/–eb* (2sg) *–o* (3sg.F/N2/N3)*–uo/–obo/–ob/–bio* (1exc/inc) *–io/–ibo/–ib* (2/3pl) (p. 262)

- "All finite verbs obligatorily have a pronominal suffix which indexes the subject" (Fedden 2011: 262).
- All features of the category (subject) are expressed in this position. There are no other position for subject indexation (see template in Appendix).
- All features are expressed by overt morphs.
- Annotate as 'yes'.

Sumerian Position 2 (Object indexation). Morphs: *–en* (1/2sg) *–Ø/–e* (3sg) *–enden* (1pl) *–enzen* (2pl) *–eš/–enê* (3pl.HUM) (Jagersma 2010: 743)

- Occurs in intransitive and transitive finite verb forms (p. 343), marking intransitive subject or transitive object. Suffixes not present in imperative (p. 343), but zero-marked imperative constructions are excluded from the investigation.
- Not all features of the category (subject/object) are expressed in this position. Position -1 also expresses transitive objects. However, these features are co-dependent: if transitive object is expressed in P-1, then subject is expressed in P2, which makes this position always filled (p. 360).
- 3sg has a zero-allomorph, which means that this position is not always phonologically filled.
- Annotate as '(yes)'.

Betoi Position 5 (mood, aspect). Morphs: *–cá* (indicative) *–idaódda/idódda* (conditional) *–dianú* (purpositive) *–ometú* (prohibitive) *–Ø* (imperative) *–ida* (imperfective) (Zamponi 2003: 21).

- Author does not mention obligatoriness of this position. Indeed, there are many examples where indicative clauses do not exhibit the indicative morph, which suggests that indicative mood is not obligatorily marked (see example 32 on p. 27 and example 20 on p. 25).
- The position is not always filled and does not express other categories than mood.
- Annotate as 'no'.

With regards to stems, the annotation considered them as part of an obligatory position. While the specific meaning does not lend itself to generalizations (because each stem expresses semantically distinct information), the morphological position prominently contrasts with other positions in terms of more schematically semantic categories: stems are the most semantically specific morphemes, and contrast with other morphs that are more general. Three languages had instances where stems were realized as zero (Ket, Burushaski and Mian). However, zero-stems are again a lexical specificity, meaning that only specific lexemes can be realized as zero stems. Because of their irregular status, zero stems do not weaken the entrenchment of the schema 'STEM–AFFIX'. An example from Ket is the verb 'to eat'.

(14) di-p-Ø
1sg-3.NA-eat
'I eat it.' (Vajda 2004: 11)

Stems might differ more than affixes in their phonological composition. As such, the obligatoriness of stems refers to merely the morphophonological dimension of the specific semantics appearing in a specific position in relation to other morphs. However, the phonological dimension of stems becomes relevant when its structure predicts the alternation of segments according to grammatical categories. Here, obligatoriness becomes a pure phonological concept. It was noted in Section 4.6.2.2 that phonological obligatoriness is difficult to determine because of the lack of an anchoring structure in which phonology and morphology can be aligned. If one defines phonological obligatoriness as obligatory filling of a morphological position, this differs from the phonological obligatoriness defined as 'a verb must have at least two syllables'. In the latter case, verbs can fill this requirements with morphs from many positions. Thus, the anchoring unit in the first case is the morphological position, in the second case the word. The cases with vowel or consonant alternation for an inflectional category have been treated as an additional position (cf. 4.4.2.3), and obligatoriness assessed on the principles outlined in Figure 4.6. For example, if final vowel alternation indexes person in some verbs, but in other verbs it merely expresses lexical meaning, this segment (final vowel) has been annotated as a separate position that is obligatory (according to the criterion that semantic obligatoriness is not necessary when phonological obligatoriness is). Infixes were, in turn, annotated with the value 'no' for obligatoriness since they can be easily considered as morphological and phonological insertions in relation to constructions where they don't appear. Stem alternation is therefore always obligatory, and infixation never obligatory. Table 4.23 shows the distribution of obligatory values across positions.

Obligatoriness Value	Yes	(Yes)	No	Total
Non-inflectional (stem)	14	3	10	27
Non-inflectional (non-stem)	0	0	159	159
Inflectional (stem)	23	4	4	31
Inflectional (non-stem)	20	18	219	257
Total	57	25	392	474

Table 4.23: Distribution of obligatory values across positions. Stem positions include stems, stem segments and infixes.

Across all samples, more than half of the 'yes' values were due to stems being always obligatory. Thus, non-stem positions are predominantly non-obligatory. In general, the criteria for obligatoriness might be too strict. However, if one looks at the obligatoriness values for inflectional morphemes alone, the discrepancy is diminished. Since non-inflectional positions are never obligatory except for stems, obligatoriness is a variable that can only show differences of morphs in inflectional positions. Chapter V will further analyze the distribution of obligatoriness across LCI, GCI and UCI positions.

4.6.3 Conclusion

The structural variables are not easy to operationalize. In order to transform the complex insights of usage-based and experimental linguistics into categories that can be detected in grammars, several strategies have been proposed: stress and tone variability are sought as categories that reflect prosodic prominence, obligatoriness of paradigms in finite verbs reflect high frequency and thus deep entrenchment. To assign a tripartite distinction of values for these two variables, a flowchart has been developed and followed in the annotation of the spreadsheet. It is expected that this operationalization of structural variables will help show the impact that prominence and high frequency have on the asymmetric distribution of complex morphology.

4.7. Conclusion

This Chapter has presented the language sample, discussed the biases that have been controlled for, and demonstrated how conditioning types, prosodic prominence and obligatoriness can be operationalized for this study. The operationalization is intended to approximate the psycholinguistic and evolutionary reality of linguistic units. The language sample provides enough data points (474 positions) to examine the correlation between the variables and types of morphs (stems, non-inflectional, inflectional). The segmentation of these positions as well as the definition of inflectional vs. non-inflectional positions entail several challenges that are related to the purpose of this study. While segmentation of verbal complexes into positions might not always reflect psycholinguistic reality, the semantic and frequency effects of morphs were taken into consideration while creating these templates. Positions account for the distinctions between individual, agglutinative morphemes, and for the semantic, morphological and phonological impact that paradigms have on the mind of speakers. The alignment of entire positions with prosody (stress, tone, weight, word-edges) and type frequency (obligatoriness) can be justified when one considers positions as having a significant effect on the mind of speakers, the more entrenched they are. Under a usage-based perspective, positions are emergent relationships between the frequent co-occurrence of morphs in a specific order. The stabilizing effect of paradigms corresponds to high type frequency. While individual morphs might exhibit different token frequencies, it is assumed that paradigms reflect a more similar frequency between their members (morphs) when they are phonologically similar.

The different flowcharts presented in the sections can be used to understand the annotation in the spreadsheet, and it is expected that results can be replicated by following them. The spreadsheet and an explanation of its abbreviations and values is found in the Appendix. While this chapter provided several tables with distributions of categories, Chapter V will closely examine the correlations between the different variables. A discussion related to the shortcomings of the methods in showing the correlations will be presented in Chapter V and VII.

CHAPTER V: COMPLEXITY, PROMINENCE, AND OBLIGATORINESS – ANALYSES AND RESULTS

5.1 Introduction

This chapter explores the relationship between the parameters, variables and values theorized in Chapter II and III and defined in Chapter IV. Two parameters are compared: inflectional complexity and stabilizing structural properties. The complexity parameter has three variables, lexically conditioned inflection (LCI), grammatically conditioned inflection (GCI) and unconditioned inflection (UCI). The category attribute ‘unconditioned’ in ‘UCI’ relates to the lack of lexical and grammatical conditioning; as such, it includes morphs that are conditioned neither by other morphs nor by grammatical features, and morphs that are only phonologically conditioned. This tripartite distinction is intended to approximate different degrees in language- and speaker-based complexity, based on information-theoretic and psycholinguistic evidence surveyed in Chapter II. However, these variables should primarily reflect the difficulty that speakers face in the acquisition and memorization of these types of morphs and the words containing them. These three types of inflection are distinguished from non-inflectional verbal categories (NI), which subsumes roots/stems without inflectional alternation, derivational and adverbial categories or categories which have not been considered to be inflectional by the criteria laid out in Chapter IV, Section 4.4.1.

The parameter ‘stabilizing structural properties’ contains two variables. The first variable is prosodic prominence, the second obligatoriness. The values for prominence are ‘yes’, ‘possible’ and ‘no’, depending on whether the morphological position in question exhibits morphs that are always, sometimes or never expressed by prominent syllables. The tripartite distinction between ‘no’, ‘possible’ and ‘yes’ should reflect an increase in prosodic prominence. The values for obligatoriness are ‘yes’, ‘(yes)’ and ‘no’, depending on whether the position is always filled by one morph, is always filled by one morph except for some zero realizations of features in the paradigm, or is not always filled by morphs. The distinction between ‘no’, ‘(yes)’ and ‘yes’ is intended to approximate an increase in type frequency of morphs and in entrenchment of morphological schemas across verbal constructions.

These two parameters are to be compared with one another. This dissertation intends to illustrate the relationship between complexity of conditioned inflection and prosodic prominence on the one hand (Section 5.2) and complexity of conditioned inflection and obligatoriness on the other hand (Section 5.3). However, the two variables from the parameter capturing structural stability will be contrasted with one another in order to see whether they are independent (Section 5.4), and whether other structural factors might explain the convergence of specific values (such as ‘yes’ values for prominence and obligatoriness).

This chapter conducts nine analyses to investigate the correlation, and significance has been determined with the help of Chi-Squared tests, one Fisher’s Exact test and a linear regression analysis. The p-value threshold has been established as 0.05, which means that results yielding a value above it will be considered non-significant. Some of these analyses are focused on the relationship between affixes and stems, and language-internal investigations are used to show conforming and deviating patterns in relation to the general correlation. The languages that conform less to the general pattern will be qualitatively analyzed and explanations for their deviation will be suggested. The presentation of results is followed by brief discussions, which are summarized in the concluding sub-sections. Section 5.5 is dedicated to a broader discussion of the relationship between the parameters, and of the benefits and shortcomings of the variables.

5.2 Complexity and prosodic prominence

Chapter III Section 3.3 concluded that certain prominent structural units are associated with facilitation and longer stability. The prominent units examined were prominent vowels and consonants, stressed syllables, syllable and word boundaries, prominent (tone) contrasts and semantically concrete morphs. Stressed syllables have been identified as the units that are not only stable but stabilize the morphology that they express. This variable has been prioritized in annotation and defined in Chapter IV, Section 4.6.1. Prominent contrasts such as tonal variability have been chosen as a variable in order to determine prominence in languages with many tones but without stress. Finally, boundaries were considered with regards to positions that only had consonantal morphs. Thus, I will only investigate these three types of prosodic prominence (hence shortly 'prominence'). Segmental prominence, such as specific vowels and consonants, as well as semantic and pragmatic prominence will be excluded from the investigation. Likewise, I will use 'complexity' to refer to the specific variable of complexity of conditioned inflection, and its values 'lexically conditioned inflection' (LCI) 'grammatically conditioned inflection' (GCI) and 'unconditioned inflection' (UCI, includes also phonologically conditioned inflection).

Before formulating the first hypothesis, I will exemplify the relationship between complexity and prominence on Ura, a moribund language of Vanuatu (Crowley 1999). Ura is one of the few Austronesian languages that exhibits LCI. Table 5.1 shows an excerpt of the annotation of this language, displaying its verbal positions with their categories, the conditioning of their inflectional morphs and prominence values.

Position	Categories and contexts of occurrence	Conditioning	Prominence
-6	Subject allomorphs for Imperative, Recent Past, Distant Past, Optative, and Future.	GCI	possible
-5	Prior past	UCI	no
-4	Negative	UCI	no
-3	Iterative	NI	no
-2	Formative; occurs with Dependent Past, Past Habitual, Present, Past Continuous.	GCI	no
-1	Derivative prefixes, reduplication	NI	no
0	Basic root used in Imperative, Recent Past, Distant Past, Optative, Dependent Past, Past Continuous, Negative, Iterative, Purposive, Instrumental, Derivational, Reduplication, Causative. Modified root used in Future, Subjunctive, Present, Habitual (pp. 148-150)	LCI	possible
1	Several object paradigms for different verbs	LCI	possible
2	Perfective, Continuative, Partitive, Misdirective	UCI	no
3	Direct/ Reflexive Object. Appears in verbs that do not have object suffixes in P1 (p. 176).	LCI	yes

Table 5.1: Verbal positions, the conditioning of their inflectional morphs and their prominence values in Ura (Austronesian)

The assignment of different conditioning values is explained as follows: P-6 morphs index subject in several paradigms dependent on tense and mood. As such, they are grammatically conditioned (GCI). Prior past *ehm(i)-* (Crowley 1999: 165) and negative *etw-/et(u)-* (p. 165) are neither conditioned by the lexicon nor by any other grammatical categories but by the phonology; as such, they received the value 'UCI'. P-2 *em(i)-/am(i)-* are formatives that co-occur with dependent past, past habitual, present, and past continuous (pp. 166-169). As determined in Chapter IV Section 4.5.2.1, formatives whose interpretation is dependent on other categories have been annotated as GCI. All other inflectional positions contain lexically conditioned morphs. Verbal stems in Ura have two shapes: a basic root (BR) and a modified root (MR). While the choice of basic vs. modified root depends on grammatical categories (p. 148-150), the formal alternation is unpredictable, and is classified into weak and strong verbs (pp. 151-155). The choice of whether verbs have object suffixes (P1), or free-standing object pronouns (P3) is also lexically conditioned: verbs that use P1 morphs do not utilize P3 morphs, and vice versa (p. 176).

With regards to the annotation of prominence, stress has been established as the prominent property in Ura: "primary stress […] is invariably found on the penultimate syllable." Since verbal roots are always (with few exceptions) polysyllabic, stress is mostly found on stem (P0) morphs:

(1) Stressed stem morph in Ura

ir^{-6}-ˈereg^{0}

3pl.PT^{-6}-BR.cry^{0}

'They cried.' (Crowley 1999: 222)

It is assumed that P1 suffixes can be also stressed when they are syllabic and precede syllabic P2 suffixes that occur at the end of the word. Crowley (1999) does not give examples of words where both P1 and P2 suffixes are present, but stresses that P2 suffixes “appear after the object suffixes [in P1] on suffixed transitive verbs, and which precede a following free-form object [in P3].” P2 suffixes are never stressed since they are always syllabic and occur at the end of words:

(2) P2 suffixes are not stressed in Ura

neveg	c^{-6}-eˈtop^{0}-ye^{2}
food	3sg:RECPT^{-6}-BR:cooked^{0}-PERF^{2}

‘The food is already cooked.’ (Crowley 1999: 178)

P3 suffixes, on the other hand, are always stressed since they are words of their own. Crowley (1999) emphasizes this fact by contrasting word forms with bound and free indexation:

(3) Stressing of P3 morphs (Crowley 1999: 174–175)

a. ci^{-6}-ˈta^{0}-qa^{1}

3sg.RECPT^{-6}-hit^{0}-2sg^{1}

‘(s)he hit you.’

b. c^{-6}-oˈcori^{0} ˈqa^{3}

3sg.RECPT^{-6}-know^{0} 2sg^{3}

'(s)he knew you.'

Example (3b) shows that two syllables can have primary stress in the verb clause 'she knew you.' One might wonder how P-6 prefixes can receive the prominence value 'possible' – these prefixes appear farthest away from the stem, and the prefixes closer to the stem are annotated with the value 'no'. The reason is that P-6 morphs, unlike other prefixes, can also be consonantal, and as such, can form the onset of a stressed stem syllable when other prefixes are not present, such as in (4):

(4) P-6 prefix forming part of a stressed syllable in Ura

ˈp^{-6}-eni^{0}

3sg.OPT-eat.BR

'(s)he ought to eat it.'[20] (161)

The other prefixes, on the other hand, are always syllabic, and as such, cannot occupy the penultimate syllable of the word since the stem is polysyllabic. The interplay of stress rules and morphological alignment (syllabicity, positions) results in a pattern where LCI positions are the most and UCI positions are the least prominent: The three LCI positions have obtained

[20] Translation is not provided by Crowley (1999: 61), but it has been attempted to translate this verb like the other examples showing the Optative on p. 161.

the values ‘yes’, ‘possible’ and ‘possible’; the two GCI positions the values ‘possible’ and ‘no’; the three UCI positions the values ‘no’, ‘no’ and ‘no’.

This association between high degree of complexity and prominence as exemplified in Ura is hypothesized to capture the greater, cross-linguistic pattern in a nutshell. The first hypothesis is thus defined as follows:

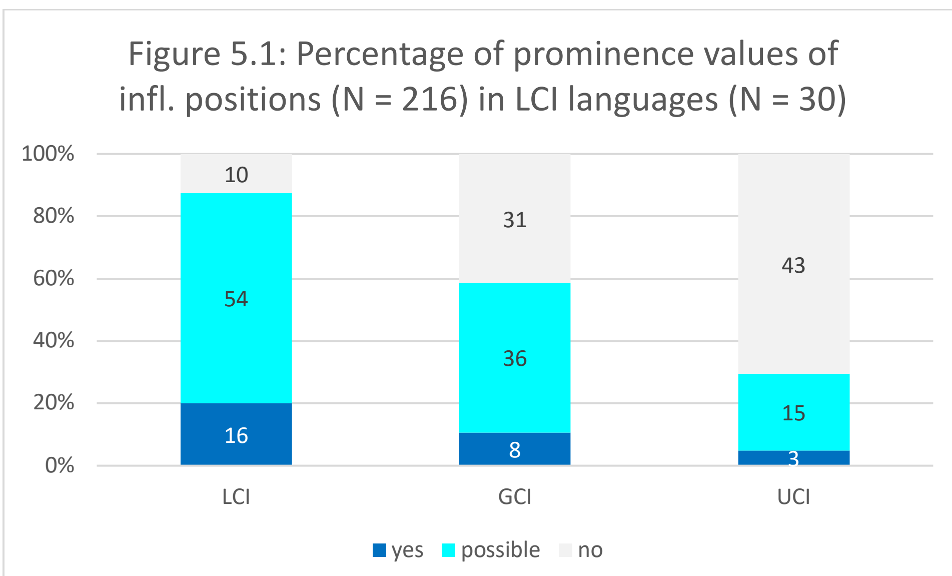
H 1.0: A higher complexity value is associated with a higher prominence value across verbal positions.

The attributes ‘high’ and ‘low’ derive from the operationalization of complexity of conditioned inflection and prominence. Thus, high complexity is manifested by LCI, middle complexity by GCI and low complexity by UCI. Highly prominent positions are the ones annotated with ‘yes’, middle prominence positions are the ones annotated with ‘possible’, and low prominence positions with ‘no’. Following Figure 4.5 in Chapter IV, consonantal morphs were assigned these values under stricter conditions to account for the fact that vowels always imply the existence of a syllable whereas consonants do not. Thus, ‘yes’ was assigned for consonantal morphs that can appear at word boundaries and are either onsets or codas of prominent syllables; ‘possible’ when they can appear at the beginning or end of words and are sometimes part of prominent syllables; and ‘no’ when they never appear in word-initial or final position, or when they are never part of a prominent syllable.

The association between the parameters ‘complexity’ and ‘prominence’ has been investigated in five quantitative analyses: the first analysis represents the general correlation (5.2.1); further analyses have been conducted to minimize other factors: the correlation of prominence with non-stem positions (5.2.2), complexity between different prominence types (5.2.3), complexity and prominence of consonantal positions (5.2.4) and proportion of inflectional and prominent positions across languages (5.2.5). Explanations for outliers from the global pattern are discussed in 5.2.6.

5.2.1 General association

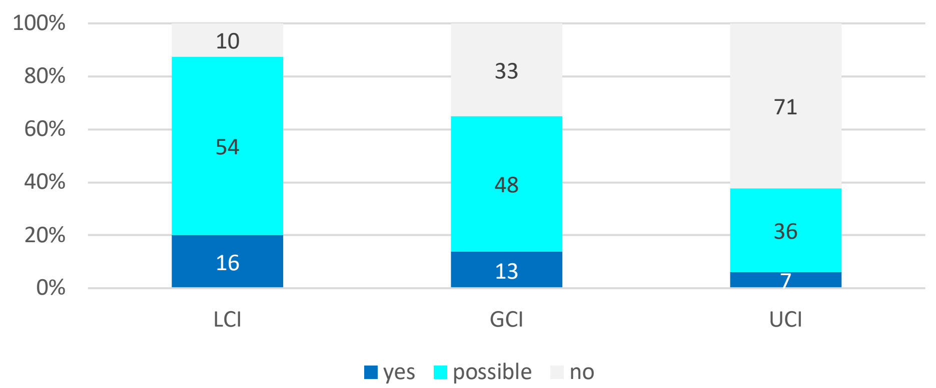
Analysis 1.0 consists of comparing the complexity and prominence values of all inflectional positions, regardless of position type (affix, stem) and prominence type (stress, tonal contrasts, word boundaries). First, the analysis was conducted on the LCI sample, which includes languages which exhibit all types of conditioning (LCI, GCI, UCI). Figure 5.1 shows the results.



LCI positions exhibit more often the prominence values 'yes' and 'possible' than GCI positions, and GCI positions more often so than UCI positions. The association shown in Figure 5.1 is highly significant ($p < 0.00001$). To calculate the significance, I used a 3x3 Chi-Squared test [values: (yes, possible, no) x (LCI, GCI, UCI)] from the web site 'Social Science Statistics'[21]. These results reflect the association of complexity and prominence that has been shown by Ura in a nutshell.

However, the question is whether the association between more complex positions (LCI and GCI) persists when adding languages to the sample that do not have LCI. As mentioned, 6 control languages that do not have LCI but GCI and UCI, and 5 languages that only have UCI have been selected. Analysis 1.0 has been run on the total sample of 41 languages, and the results are shown in Figure 5.2.

Figure 5.2: Percentage and number of prominence values of positions (N = 288) in all languages (N=41)

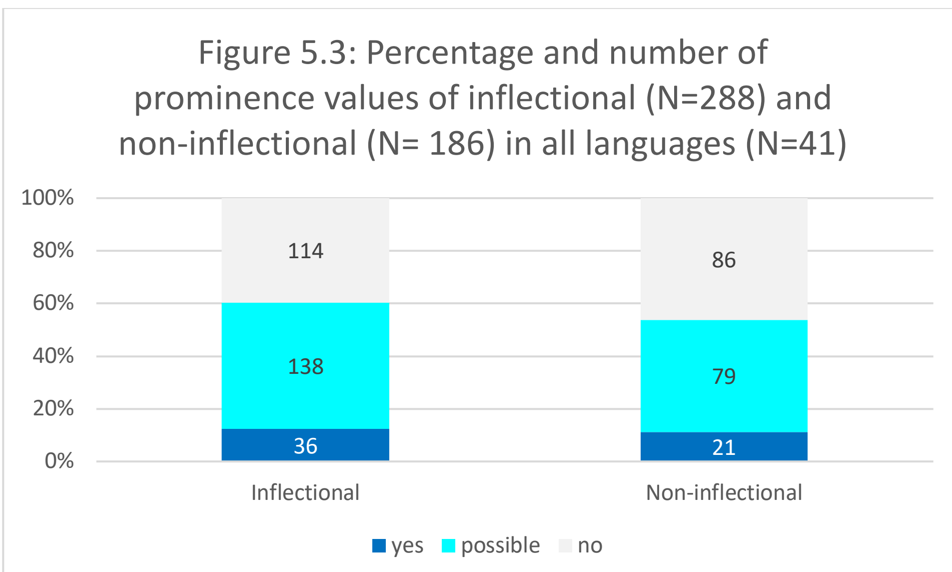


Here, the association between more complex and more prominent positions is not changed once the data from the control samples is incorporated and is also highly significant ($p < 0.00001$). There is a considerable rise in GCI positions that are associated with 'yes' and

[21] https://www.socscistatistics.com/tests/chisquare2/default2.aspx

‘possible’ values compared to Figure 5.1. This shows that within GCI languages that do not have LCI, GCI tends to occupy environments that are more prominent in comparison to UCI. Nevertheless, the association between prominence and LCI in LCI languages outweighs this effect observed in languages lacking LCI.

To show that the association between prominence and positions is dependent on conditioning of inflection, and not on inflection in general, inflectional positions and non-inflectional positions have been compared across all samples. Figure 5.3 shows that the difference between the association of prominence with inflectional categories and the association of prominence with non-inflectional categories (non-alternating lexical, derivational and adverbial morphs) is very small. In fact, the difference is not statistically significant (p=.357969).



Quite interestingly, the lack of association between prominence and inflectional or non-inflectional positions remains in languages that have only UCI (N=6). In these languages, inflectional positions (N=25) align at a rate of 48% with ‘yes’ or ‘possible’ values and non-inflectional positions (N=25) at a rate of 56% with those values. This is a lower value for inflectional positions compared to Figure 5.3, where they align with prominence values ‘yes’ and ‘possible’ at a rate of 60% compared to non-inflectional positions (54%); however, this may be explained by the fact that UCI is generally associated with more ‘no’-values, even if it is the only type of inflection in languages.

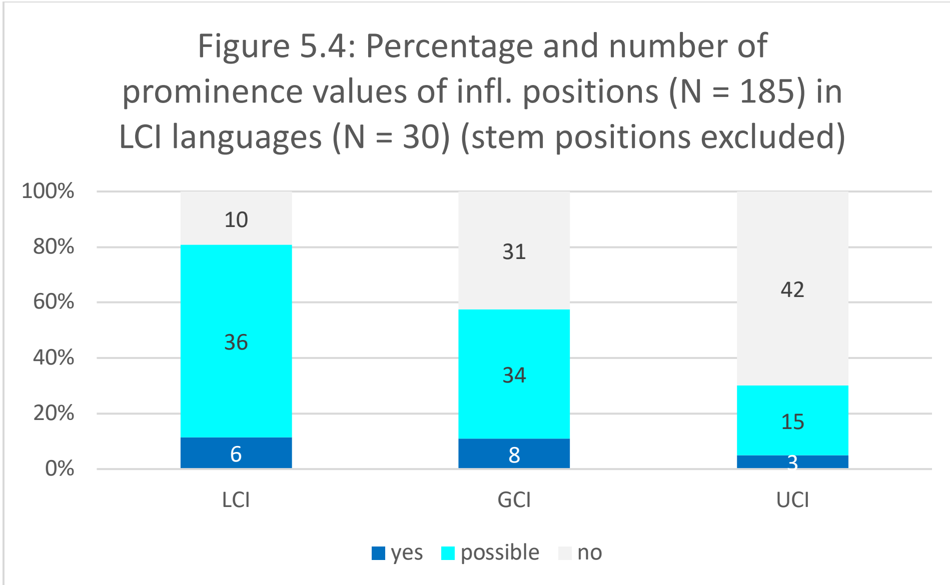
Analysis 1.0 conducted on the LCI languages and across all samples confirms Hypothesis H1.0: LCI is associated with a higher prominent value than GCI, and GCI is associated with a higher prominent value than UCI. GCI and UCI behave similarly in languages with and without LCI. Furthermore, it is rather the more complex inflection types that pattern with prominent positions and not inflection in general in comparison with non-inflectional positions. In 5.2.2., the non-stem positions will be analyzed, to control for the effect that semantic and morphological prominence of the stem has on the distribution of complexity.

5.2.2 Non-stem positions

In addition to analyzing the general association between complexity and prominence types, non-stem positions were also analyzed. Chapter III, Section 3.3.2.1, stated that stems are prominent because they stand out from other affixes by being semantically more concrete than affixes. Roots or stems are considered a ‘faithful position’ according Beckman (1998) which is less susceptible to morphological or phonological changes. Furthermore, Beckman (1998: 191-192) states that stems are more likely to attract stress than affixes, and this makes sense historically since semantically or pragmatically prominent constituents often attract prosodically prominent features (Frota 2014; Boye and Harder 2012; Büring 2010; Gundel 1988; Samek-Lodovici 2005; Selkirk 1995; Szendroï 2003). Second, inflectional morphs expressed by stem alternation are almost always lexically conditioned, that is, the alternation is not predictable from the form or meaning of the stem. Thus, it could be the case that many of the instances where prominence and LCI converge is due to stems attracting both LCI and prominence. However, given that there is more evidence for stabilizing effects of stressed syllables as opposed to semantically more concrete units (see Table 3.3 in Chapter III), one could still expect that LCI morphs are more likely expressed by stressed syllables even in non-stem positions. To investigate this, I formulated the sub-hypothesis H 1.1:

H 1.1: A higher complexity value is associated with a higher prominence value across verbal non-stem positions.

The following charts present the results obtained after excluding stem morphs. Stem positions are those positions which authors of the grammars claim to be the stem or root. Thus, for example, Betoi verbs can have two stem elements, separated by subject interfixes (Zamponi 2003: 23). Because stem positions were excluded in order to control for potential prominence due to semantic concreteness, recurring incorporated elements or lexically conditioned auxiliaries have not been considered as stems, as they tend to have more general (and thus less concrete) semantics. As mentioned in Chapter IV Section 4.4.2.3, stems were split up into additional positions if there was a consistent alignment between morphology and phonology, such as the three-consonantal pattern in Afroasiatic languages, as is the case in Zuwara Berber. Each segment associated with specific lexical meaning has been therefore excluded from Analysis 1.1. Figure 5.4 shows the results for Analysis 1.1 with regards to the LCI sample.



The results of Analysis 1.1 show a similar pattern to the ones of Analysis 1.0. LCI attracts more prominent values ('yes'/ 'possible') than GCI, and GCI more than UCI. The only difference is within 'yes' values alone, where the difference is minimal across inflectional types. However, the Chi-Squared test between prominence and more complex inflectional non-stem positions is again highly significant ($p < .00001$). Most of the excluded positions were indeed LCI (28 out of 80), however, despite this, the association of prominence and inflectional complexity still holds with regards to general, less specific morphs (i.e., affixes). That is, LCI attracts more prominence regardless of whether it is expressed by stem alternation. Across all samples, we see a similar pattern (Figure 5.5).

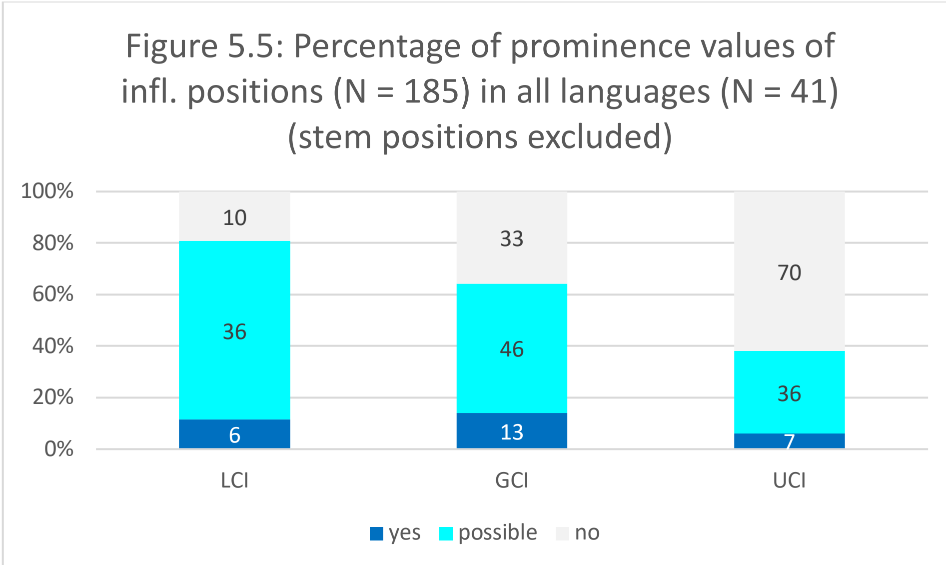


Figure 5.5 shows that the tendency for LCI to attract more prominent positions than GCI and UCI also holds when incorporating the two control samples. This association is also statistically significant ($p < 0.00001$). However, now the difference between the 'yes'-values becomes more evident: GCI positions attract more often always prominent syllables than LCI. Since GCI, unlike LCI, is rarely expressed through stem alternation, excluding stems that always attract stress in some languages affects the distribution between always prominent positions and conditioning type. Nevertheless, the sub-hypothesis H 1.1 can be confirmed: a higher complexity value ('yes' and 'possible') is associated with a higher prominence value (LCI and GCI) across verbal non-stem positions, as is the case with stem positions included.

5.2.3 Prominence types

A third analysis has been conducted on the distribution of the two major prominence types for positions with syllabic morphs. As determined in Chapter IV Section 4.6.1.5, stress was prioritized in the annotation of prominence. Thus, if an author mentioned that the language had stress, this was always taken as a sign of prominence. If authors did not mention that the language had stress, but associated specific tones with prominence, then the presence of these tones on morphs determined the prominence values assigned to their positions. Finally, if authors indicated neither stress nor specific prominence for tones, high tonal variability was considered as an indicator for prominence. Furthermore, if a language did not have stress, and only three or fewer tones, prominence was associated with the highest tone if that tone was not the most frequent one. Finally, if a language had three or fewer tones, and the highest tone did not appear to be the least frequent one (i.e., marked tone), heavy syllables have been categorized as prominent. This last criterion was only used for one language, Bangime.[22] Table 5.2 shows the properties used to determine prominence across languages.

[22] Syllable weight might be a general indicator of prominence, but due to the complication in determining weight across languages, weight was not analyzed as a general variable.

Prominence based on	Languages selected
Stress/prominent syllables	Aguaruna, Ayutla Mixe, Azari, Bardi, Betoi, Burushaski, Dumi, Estonian, French, Hatam, Ingush, K'ichee', Koasati, Korean, Luo, Mapudungun, Murrinh-Patha, Navajo, Northern Pomo, Nunggubuyu, Nyamwezi, Oneida, Pilagá, Semelai, Sumerian, Supyire, Yurok, Tunisian Zuwara Berber, Ulwa, Ura, Yelî Dnye, Yurok
Prominent tones	Ket, Kiowa, Lumun, Mian, Nyamwezi
High tonal variability	ǂHȍã, Sheko, Zacatepec Chatino
Syllable weight	Bangime

Table 5.2: Prominence types across languages of all samples

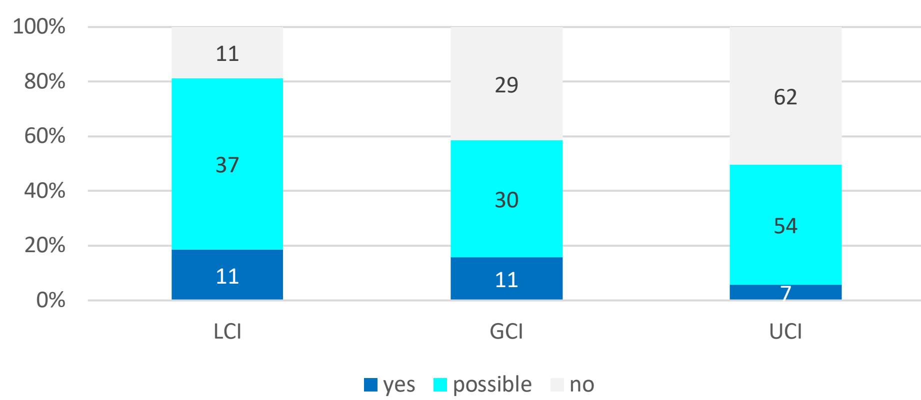
Table 5.2 shows that most of the languages have stress, and in five other languages, specific tones could be identified as prominent either by the authors' mention, or by assuming the highest tone is prominent if it is the marked (least frequent) tone in languages with two or three tonal contrasts. For ǂHȍã, Zacatepec Chatino and Sheko, tone variability was the determining factor for assigning prominence values to positions. 'yes' was assigned to positions that exhibit morphs that can have any tone from the tonal inventory, 'no' to positions whose morphs can only have one tone, and 'possible' for positions whose morphs cannot have all the tones but more than one tone. Finally, in one language, Bangime, heavy syllables were used as the prominence-defining property since the language does not show evidence for stress nor for specific marked tones.

In order to justify the subsumption of all prominence types into three values of prominence, it must be assumed that the relationship between prominence and complexity does not differ among prominence types. Thus, another hypothesis is needed to prove this relationship across prominence types:

H 1.2: The relation between inflectional type and prominence should be similar within each inflectional type and across all types.

Analysis 1.2 has been run on all samples. Figures 5.6–5.9 contrast the distribution of prominence values across the types of prominence listed in Table 5.2.

Figure 5.6: Percentage and number of prominence values of infl. positions (252) among languages with stress (N = 32)



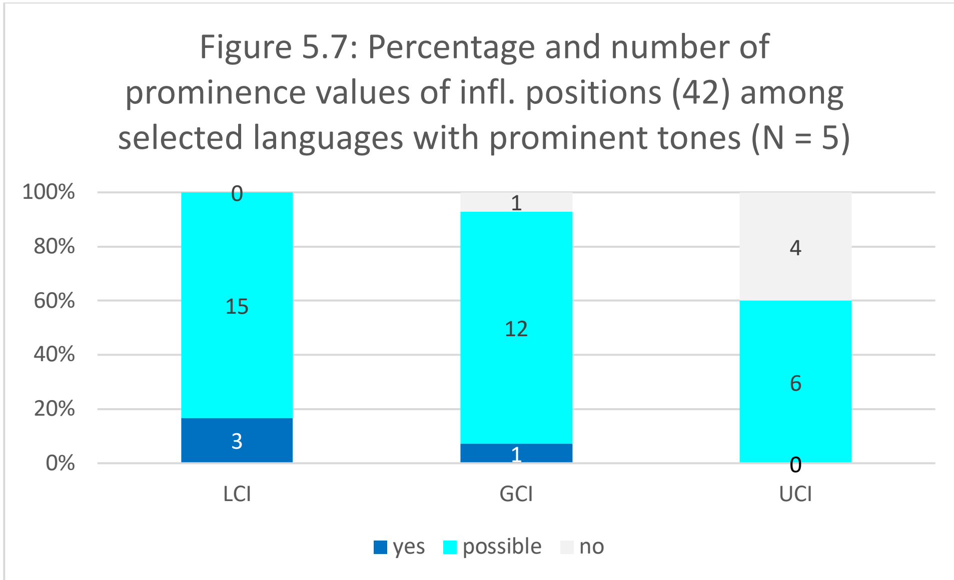


Figure 5.7: Percentage and number of prominence values of infl. positions (42) among selected languages with prominent tones (N = 5)
100%
80%
60%
40%
20%
0%
0
15
3
1
12
1
4
6
0
LCI
GCI
UCI
yes
possible
no

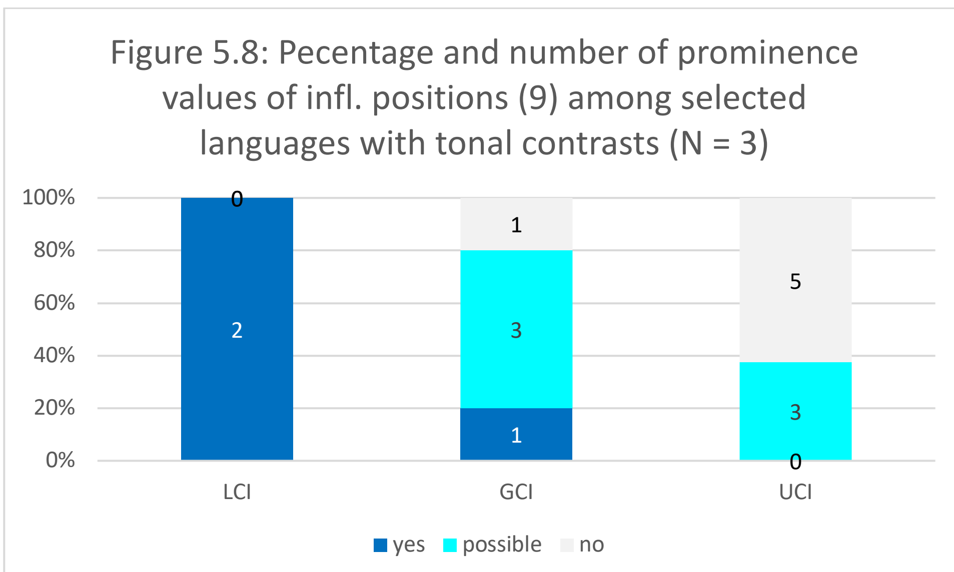
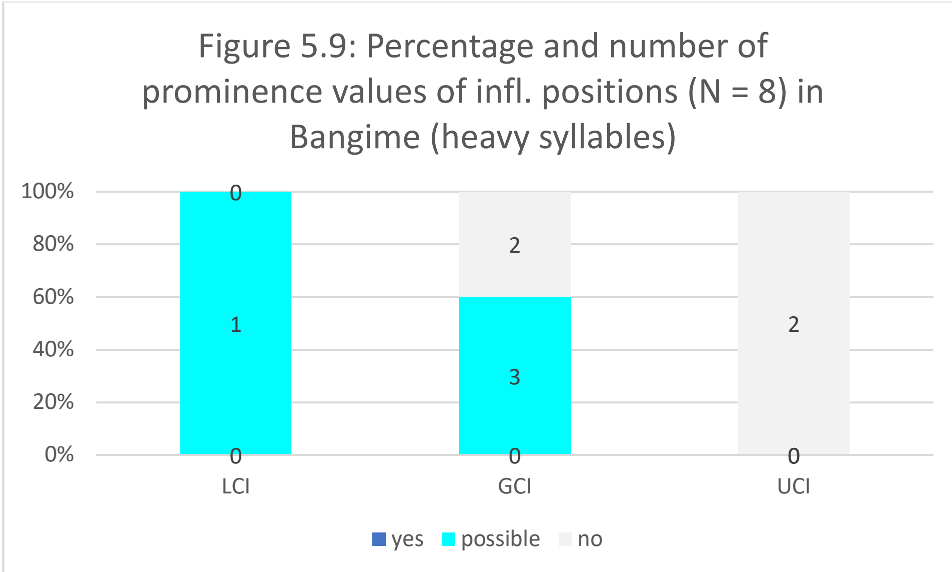


Figure 5.8: Pecentage and number of prominence values of infl. positions (9) among selected languages with tonal contrasts (N = 3)
100%
80%
60%
40%
20%
0%
0
2
1
3
1
5
3
0
LCI
GCI
UCI
yes
possible
no



Figures 5.6–5.9 show that the distribution of morph types is similar across prominence types. However, due to the scarcity of occurrences of morph types (including empty cells for some prominence values) for the last three prominence types (Figures 5.7–5.9), stress is the only prominence type for which the Chi-Squared test could be calculated. This association is significant (p =. 000357). UCI is associated more with stress than with any other type of prominence. Indeed, all of the 'yes' values for UCI are contributed by languages with stress. On the other hand, languages where prominence has been determined based on tonal variability exhibit most of their tonal contrasts in stems, which is the locus of lexically conditioned inflection. In Zacatepec Chatino and Sheko, stem alternation is also associated with tonal alternation, which explains the alignment of prominence with LCI. Despite the less numerous prominence types yielding non-significant results due to the lack of data for certain types (in 78 % of all languages, prominence has been determined based on stress), the results of Figures 5.7–5.9 show a trend into the same direction as the results from the overall analysis: positions considered more prominent than others are associated with more complex inflection.

H 1.2. can be confirmed: there is a statistically significant association between stress and higher inflectional complexity, and while significance was not determined for the alignment of the other types of prominence with complexity, they show a pattern that behaves according to the overall analysis. Interestingly, whenever UCI is associated with prominent positions, it is stress that determines this prominence. This association could be interpreted in terms of stress languages frequently anchoring main stress around word edges, which is the place where UCI tends to occur. Tonal contrasts, on the other hand, are more likely instantiated in stem positions, and stems morphs occur least often at word boundaries, given the existence of prefixes and suffixes.

5.2.4 Consonantal positions

Finally, another analysis has been conducted to investigate the behavior of the phonotactic type 'consonantal morphs' whose prominence has been assessed following stricter criteria. Because stress and tone are always associated with vowels and syllabic consonants, the prominence of positions that only exhibited (non-syllabic) consonantal morphs could not be evaluated in a straightforward way. In order to ensure that the consonants themselves are likely to be prominent, two criteria were applied to determine prominence values, namely potential to occur word-initially (as an onset) or word-finally (as a coda), and the potential to be either onset or coda of a prominent syllable. The criterion for word position was ranked higher than the criterion for belonging to a prominent syllable, as outlined in Figure 4.5 in Chapter IV – consonantal morphs that never occurred at the edge of a word but were part of prominent syllables received the value 'no' for prominence. On the other hand, word-edge was not considered independently a criterion for prominence, since only word-initial or word-final consonantal morphs that also occurred in stressed syllables were considered prominent. Analysis 1.3 tests whether consonantal morphs from all samples show a different behavior in comparison to the general association presented in Figures 5.1–5.2. Because of the stricter criteria, the assumption is that there should be no difference. Hypothesis H 1.3. is formulated as follows:

H 1.3: A higher complexity value is associated with a higher prominence value across positions that only have consonantal morphs.

21 purely consonantal inflectional positions were investigated in relation to the prominence values ‘yes’ ‘possible’ and ‘no’. Figure 5.10 shows the results.

Figure 5.10: Percentage and number of prominence values of positions with purely consonantal (non-syllabic) infl. morphs (N = 21; all samples).

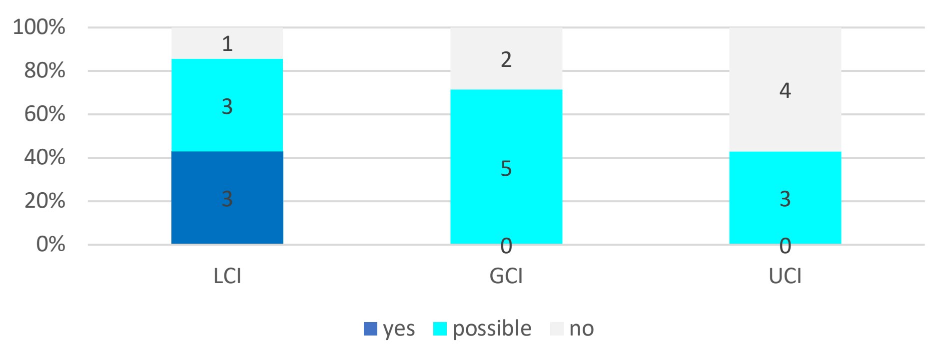


Figure 5.11 shows a distribution that is quite similar to the global pattern presented in Figures 5.1 and 5.2. However, due to the scarcity of data, a Chi-Squared test could not be performed. A 2x3 Fisher Exact test where ‘yes’ and ‘possible’ were grouped together to avoid zero values yielded non-significant results (P_A= 0.381; P_B= 0.159); As such, H 1.3 cannot be

confirmed, but also not refuted. However, again, the trends are in the direction of the global pattern: consonantal LCI patterns more often with prominent positions than consonantal GCI, and GCI more so than UCI.

5.2.5 Proportion of inflectional and prominent positions across languages

When annotating prominence across templates in languages, the following pattern was observed: Languages that have multiple positions with morphs of high complexity (LCI and GCI) have a more variable location of prominent syllables with regards to morphological positions. Languages with only one position of high complexity are more likely to align prominence with this position. An example of a language with multiple complex positions is Ket. In Ket, almost every position contains lexically conditioned inflection. The rising-falling tone contour, which is considered prominent (Vajda 2003: 20), can align with every morphological position, as has been shown in Denk (2020). Table 5.3 presents an excerpt of the annotation in Ket, and examples (5 a-e) show the alignment of the contour with these positions.

Position	Categories/Functions	Conditioning	Prominence
-8	Subject, involuntary causa-tive	LCI	possible
-7	Incorporate	NI	yes
-6	Subject/ object	LCI	possible
-5	Thematic, adverbial	NI	possible (con-sonantal)
-4	Object, present, past, re-sultative	LCI	possible
-3	Subject/ object, applicative, pluractionality	LCI	Possible (con-sonantal)
-2	Past, imperative	LCI	possible
-1	Subject, object	LCI	possible
0	Root, transitivity, iterative, semelfactive	NI	possible
1	Subject	LCI	possible

Table 5.3: Verbal positions, conditioning and prominence of morphs in Ket (Yeniseian)

(5) Alignment of rising (´) and falling tone (`) syllables with morphological positions in Ket (position number is indexed by superscript) (Vajda 2004).

a. **dí$^{-8}$-ròq^{0}**

1.S^{-8}-fly^{0}

‘I fly.’ (p. 56)

b. da^{-8}-**bú$^{-6}$-g^{-5}-dì$^{-1}$**-qos^{0}

3.F/N.A^{-8}-**3sg.A^{-6}-TH^{-5}-1sg.O^{-1}**-bring^{0}

‘She brings me.’ (p. 54)

c. d^{-8}-**ó$^{-4}$-l^{-2}-Ø^{0}-ìn^{1}**

3AN^{-8}-**3sg^{-4}-PT^{-2}-live^{0}**-pl^{1}

‘They lived.’ (p. 92)

d. d^{-8}-**b^{-3}-ín^{-2}-tèt^{0}**

3.M^{-8}-**3.N^{-3}-PT**$^{-2}$-hit^{0}

‘He hit it.’ (p. 50)

e. **dón^{-7}-ìl^{-2}**-da^{-1}-bet^{0}

knife^{-7}-PT^{-2}-1pl^{-1}-have^{0}

‘I have a knife.’ (p. 49)

As one can see, every morphological position (P-8 – P1) can be associated with a rising (´) or falling (`) tone.

On the other hand, there are languages where complexity of conditioned inflection is confined to a few positions. Bardi represents an extreme case where only one position has GCI morphs, whereas the remaining eight inflectional positions contain only UCI morphs. Stress falls only on the position with GCI morphs. The template of Bardi is presented in Table 5.4.

Position	Categories/Functions	Conditioning	Prominence
-6	Subject	GCI	yes
-5	Transitivity (clitic)	NI	no
-4	Past, present, future, irrealis	UCI	no
-3	Subject	UCI	no
-2	Transitivity (clitic)	NI	no
-1	Pluractionality	NI	no
0	Root	NI	no
1	Applicative	NI	no
2	Reflexive	NI	no
3	Continuative, remote past	UCI	no
5	Applicative	NI	no
6	Recent completed past, middle perfect, future	UCI	no
7	Simultaneous, subordinator	NI	no (consonantal)
8	Linker, relator, contrastive, re-sumptive	NI	no
9	Topic	UCI	no
10	Oblique	UCI	no
11	Possessor	UCI	no
12	Direct object	UCI	no
13	Quantifier	NI	no

Table 5.4: Verbal positions, conditioning and prominence of morphs in Bardi (Nyulnyulan, Australia)

The fact that only the morphs in P-6 are stressed derives from the fact that this position is obligatory because subject/tense marking is obligatory, and is therefore always filled by syllabic units, namely *nga* (1) *mi* (2.present/past) *a* (2.bivalent.future/irrealis/imperative) *nga* (2.monovalent.future/irrealis/imperative) *i* (3.past/present) *oo* (3.future/irrealis) *a* (1+2; 1.aug) *goo* (2.aug) (180-181) *ma* (gerund/infinitives) (Bowern 2012: 396, 452) and the fact that "primary stress is regular and appears consistently on the initial syllable of the word" (Bowern 2012: 110). Some examples are given in the following:

(6) Stress placement in Bardi

a. ˈ**i**$^{\mathbf{-6}}$-ŋanka^{0}-n^{3}-j^{7}

3(PR/PT)$^{\mathbf{-6}}$-speak^{0}-CONT^{3}-SIMUL^{7}

'while he's talking' (Bowern 2012: 116)

b.

ˈara	ˈ**ŋa**$^{\mathbf{-6}}$-l^{-4}-iɲa^{0}-n^{3}	ˈaːɭi
NEG	**1**$^{\mathbf{-6}}$-IRR^{-4}-catch^{0}-CONT^{3}	fish

'I didn't catch a fish.' (Bowern 2012: 117)

The behavior in Ket and Bardi suggests that there could also be a correlation between the number of inflectional positions with high complexity and the number of positions that are or can be prominent across verbal constructions. This seems plausible following the theory that prominent syllables stabilize morphology. If prominence is restricted to a specific part of the word, the complexity accrued there is less likely to erode phonologically or undergo morphological analogy. If morphs from many positions can be prominent, morphological complexity is more likely to be retained across positions. Thus, a further analysis that investigates this relationship has been conducted, with the goal to support the following hypothesis:

H 1.4: A higher percentage of complex inflectional positions (LCI, GCI) in a language is associated with a higher percentage of positions that receive prominence (‘yes’, ‘possible’).

In order to calculate the association between the percentage of complex positions in languages (dependent variable) and the percentage of prominent positions (independent variable), a linear regression was tested. Complex positions were defined as the ones annotated as LCI and GCI, and the proportion was determined in relation to the sum of all positions (that is, UCI and NI were considered non-complex). With regards to prominence, the proportion of prominent positions resulted from the ratio of positions that have ‘yes’ and ‘possible’ values in relation to all positions (that is, ‘yes’ and ‘possible’ values were considered prominent, and ‘no’-values non-prominent). Analysis 1.4 yields a significant linear regression ($p = .018$; standard deviation of residuals: 0.241)[23] between these two proportions, showing a positive slope.

[23] I used a regression ANOVA from the web site ‘Statistics Kingdom’ (https://www.statskingdom.com/linear-regression-calculator.html) to determine the P-value of the linear regression. To illustrate the regression in Figure 5.11, I used Microsoft Excel.

Figure 5.11: Proportion of complex inflectional and prominent positions in LCI and GCI languages

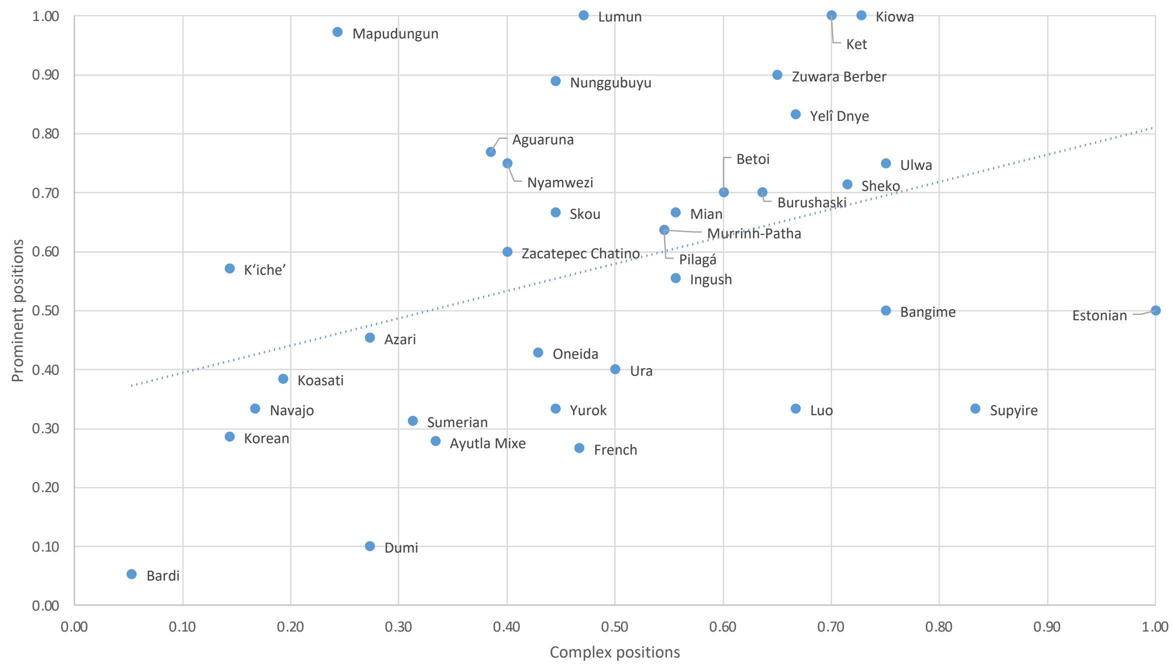


Figure 5.11 shows that there is a tendency for languages with a larger proportion of complex positions to have a larger proportion of positions that can be prominent, even after expressing complexity and prominence potential in binary terms. Thus, not only is inflectional complexity associated with prominence on the paradigmatic dimension, but also on the syntagmatic dimension. This suggests that languages with verbs that concatenate many complex morphs tend to align these morphs with prominent syllables. Languages that diverge from this trend are, for example, Mapudungun and Estonian. This can be explained by the number of positions they have. Mapudungun represents the highest number of positions in the sample (37), Estonian the lowest (2). In Mapudungun, 36 positions can be stressed, but only 9 are annotated as complex (GCI). In Estonian, both positions can exhibit LCI, but stress always falls on the first syllable, which is the stem. Thus, in Mapudungun, the high number of positions leads to an overrepresentation of instances where non-complex positions, even with only one or few morphs, are

associated with prominence. In Estonian the fact that both positions exhibit LCI results in a 50–50 distribution of prominent vs. non-prominent complex positions.

It should be reminded that Figure 5.11 does not consider the number of positions, only the proportion of complexity and prominence. Furthermore, Figure 5.11 does not show how values are distributed across the template, nor whether the distribution of complexity and prominence align. For example, in Ket, all 10 positions exhibit LCI and are potentially prominent. On the other hand, in Korean one position out of seven is complex (GCI), and two are potentially prominent, but these positions do not contain GCI. The complex position is the last position (P5), whereas the potentially prominent positions are at the beginning of the template (P-1, P0). Thus, in order to see how languages deviate in their alignment between complexity and prominence, Analysis 1.0 (which investigated whether a higher complexity value is associated with a higher prominence value across verbal positions) is a better indicator than Analysis 1.4, but Analysis 1.0 does not provide information about the intra-linguistic relationship between complexity and prominence or about how well a language aligns with the global pattern.

5.2.6 Explaining outliers in the LCI sample

This section presents examples that contradict the general pattern. The relevant exceptions are where high complexity is expressed by non-prominent syllables, and low complexity by prominent syllables – the cases where LCI positions received the prominence value ‘no’ and UCI positions that received the value ‘yes.’ I will only focus on instances from the LCI sample, since there is no UCI position with the value ‘yes’ in the GCI-sample, and UCI receiving ‘yes’ values in the UCI sample is not as surprising given the absence of more complex types of inflection. Table 5.5 and 5.6 present the positions and (some of) the morphs that defy the global pattern.

Language	Position	Morphs and function
Navajo	-2	Different conjugations for subject depending on aspect, e.g. ni-imperfective: *nish* (1sg) *ní* (2sg) *(y)í* (3sg) *nii(d)* (1du/pl) *no(h)* (2du/pl) *aa* (3pl) *í* (3.nonspecific) *jí* (3.nonpersonal) si-perfective: *sé* (1sg) *síní* (2sg) *(yi)s* (3sg) *sii(d)* (1pl) *soo* (2pl) *aaz* (3pl) *iz* (3.nonspecific) *jiz* (3.impersonal) (...) (see all paradigms in Young and Morgan and Midgette (1992: 907-921), also in combination with other affixes (P-8 – P-3)
Dumi	2	*k* (1pl) *n/Ø* (1sg > 2) *ŋ* (1sg) (Van Driem 1993: 121)
Supyire	1	*lì* (imperfective) (130) *ni* (imperfective) (134) *gè* (134) *Ø* (imperfective) (e.g., on p. 135 in variant alternation with *ni*, otherwise when there is only root-internal alternation like in p. 133) *re* (imperfective) (136) different last vowel of the stem (e.g., *tuugo* 'accompany (perfective)', *tuuge* 'accompany (imperfective)', p. 132, *bubo* 'not be well shut (imperfective)', *bùbi* 'not be well shut (perfective)', p. 133). *gV* (causative in some verbs) (142) *lV* (non-productive iterative/intensive/participant plurality) (Carlson 1994: 145)
Pilagá	-1	Set A: *s* (1) *aw/o* (2) *d/t/i/yi/h/Ø/w* (3) Set B: *ñ* (1) *an* (2) *n* (3) (Vidal 2001: 136)
Nunggubuyu	3	*ny* (past/nonpast 1) *ngi* (past2, nonpast3) *ng* (nonpast 1, past 1) *na* (nonpast 2) *ni* (nonpast 2, past 2) *Ø* (nonpast 3, past 2, nonpast 2) *ngun/ngan/nyji* (evitative) *di* (past 2) *ji:* (nonpast 2) *jan/n* (evitative) *ra* (nonpast 2) *y* (past 2) (Heath 1984: 408-411).
Estonian	1	Present/Indicative: *n* (1sg) *d* (2sg) *b* (3sg) *me* (1pl) *te* (2pl) *vad* (3pl); Present Conditional: *ksin* (1sg) *ksid* (2sg) *ks* (3sg) *ksime* (1pl) *ksite* (2pl) *ksid* (3pl); Past/Imperfect: *sin* (1sg) *sid* (2sg) *s* (3sg) *sime* (1pl) *site* (2pl) *sid* (3pl) (252); *da/ta/a* (infinitive) *ge/ke* (2pl imperative) *gu/ku* (jussive) *nud* (past participle) *ma* (supine) *v* (present participle) *vat* (evidential) *da/ta* (impersonal negative indicative) *dakse/takse/akse* (impersonal present) *di/ti* (impersonal past) *dud/tud* (impersonal past participle) *gu/tagu/dagu* impersonal imperative) (Blevins 2007: 252)
Tunisian Zuwara Berber	1	*V* (present, conjugation 11) (Mitchell 2009: 27-31)
Yurok	2	E-Class: *Cek'* (1sg) *Ce'm* (2sg) *C'/'C* (3sg) *Coh* (1pl) *Cu'* (2pl) *Ceł* (3pl) (34) *es* (imperative singular) (Robins 1958: 44)

Table 5.5: LCI positions with the value 'no' for prominence (LCI-sample).

Language	Posi-tion	Morphs and function
Skou	2	*ka* (negation) (Donohue 2004: 264)
Koasati	5	*áhi* (intention) *á* (immediate intent) (Kimball 1991: 159)
Betoi	4	*omé/óme* (negation) (Zamponi 2003: 34)

Table 5.6: UCI positions with the value 'yes' for prominence (LCI-sample).

There are two ways one can approach exceptions. The first approach is to use them to justify non-universal behaviors across data, and to weaken universal claims. The second is to use them to justify that the universal is still present, but one must look at the larger context in which these exceptions occur. Because this dissertation aims to find universals, the second approach will be chosen. Thus, the contexts of the exceptions will be further analyzed to justify the universal tendency. Since there are only a few exceptions, this justification must be done qualitatively.

5.2.6.1 Non-prominent LCI positions

With regards to LCI, the question is which other factors can explain its distribution in positions that received the value 'no' for prominence. One factor might be related to the fact that not all types of prosodic prominence have been annotated. Syllabic morphs are annotated as prominent or non-prominent merely based on their behavior regarding stress and tonal properties. Because prosodic boundaries are an additional prominent property independent from stress, one could expect that LCI which does not receive syllabic prominence might have positional prominence, that is, it would appear often at the beginning or end of words.[24] A second factor will be investigated in Section 5.3, namely obligatoriness. According to the arguments presented in Chapter III, obligatoriness could explain why complex inflection persists among languages, independently from prominence. Thus, one could expect that LCI that is never prominent might survive in languages because of its high type frequency. Third, non-prominent LCI positions might occur in verbs where there are other LCI positions that in their majority already align with prominent positions. That is, a language might not be able to accommodate all LCI positions as prominent, when prominence is restricted to specific parts of the word (however, the case of Ket shows that all LCI positions can align with prominent syllables). A fourth factor that might account for non-prominent LCI positions has to do with the definition of LCI itself. Non-prominent LCI positions might have received the value 'LCI' according to weak criteria. Strong LCI positions would be the positions were every morph is lexically conditioned, whereas weak LCI positions have LCI morphs in their minority. Parallel to that, the values for prominence might not reflect strong prominence. For example, in a language where stress falls on every second syllable, there might not be a strong association between a specific stressed syllable and a morphological position, reducing the distinctiveness of stress as a contrasting feature. In the following, I have investigated whether one or more of these factors can explain the exceptions in favor of the universal trend. Table 5.7 summarizes whether the abovementioned properties are present in the LCI positions that received the prominence value 'no.'

[24] On the other hand, the relationship between prosodic prominence of boundaries and stability of word boundaries is more complicated, as discussed in Chapter III Section 3.3.1.3. In the following, only word-initial and word-final positions will be analyzed, even if 'word-internal onsets after obstruent codas' are the strongest positions according to Ségéral and Scheer (2008).

Languages and LCI positions		Appears at word edge	Obliga-tory	Other LCI positions present that can be prominent	Weak LCI status	Weak prominence status	Un-weighted total property value
Navajo	**P-2**	**Possible**	**Yes**	**Yes**	**No**	**Yes**	**3.5**
	P0	Possible	Yes	No	No	No	1.5
Dumi	**P0**	Possible	Yes	No	No	No	1.5
	P2	**No**	**No**	**Yes**	**No**	**Yes**	**2.0**
Supyire	**P0**	Possible	Yes	No	No	No	1.5
	P1	**Yes**	**No**	**Yes**	**No**	**No**	**2.0**
Pilagá	**P-3**	Possible	No	Yes	No	No	1.5
	P-1	**Possible**	**Yes**	**Yes**	**Yes**	**No**	**3.5**
	P0	Possible	No	Yes	No	No	1.0
	P2	Possible	No	Yes	No	No	1.0
	P3	Possible	(Yes)	Yes	No	No	2.0
	P6	Yes	No	Yes	No	No	2.0
Nunggubuyu	**P0**	Possible	Yes	Yes	No	Yes	3.5
	P1	Possible	No	Yes	No	Yes	2.5
	P3	**Yes**	**No**	**Yes**	**No**	**Yes**	**3.0**
Estonian	**P0**	Yes	Yes	No	No	No	2.0
	P1	**Yes**	**Yes**	**Yes**	**No**	**No**	**3.0**
Tunisian Zuwara Berber	**P-8**	Possible	No	Yes	No	No	1.5
	P-2	Possible	No	Yes	No	No	1.5
	P0.1 (first stem consonant)	Possible	Yes	Yes	No	No	2.5
	P0.2 (stem vowel)	No	No	Yes	No	No	1.0
	P0.3 (second stem consonant)	Possible	(Yes)	Yes	No	No	2.5
	P0.4 (stem vowel)	Possible	No	Yes	No	No	1.5
	P1 (present)	**Possible**	**No**	**Yes**	**No**	**No**	**1.5**
	P1 (negation)	Possible	No	Yes	No	No	1.5
	P5	Yes	No	Yes	No	No	2.0
Yurok P2	**P2.1 (o-class morphs)**	Yes	(Yes)	No	Yes	No	3.0
	P2.2 (e-class morphs)	Yes	(Yes)	Yes	Yes	No	4.0

Table 5.7: Presence of further properties of LCI positions with the prominence value ‘no’ (white) and ‘yes/possible’ (grey).

The anomalies in the languages are contrasted to the LCI positions that conform to the global pattern (i.e., receiving ‘yes’ or ‘possible’ values). Presence of a property has been operationalized (No = 0; Possible = 0.5; Yes/(Yes) = 1) to account for differences in the number of properties pertaining to the LCI positions. As Table 5.7 shows, LCI positions that receive the prominence value ‘no’ also exhibit other properties that might explain their exceptional status. Moreover, most of the conforming positions have a lower unweighted total property value than the non-conforming positions.

In Navajo, P-2 affixes express cumulatively tense, mood, aspect and person, and can appear at the left word edge, are obligatory and are not the only LCI position in the template. Furthermore, Navajo exhibits a non-canonical marking of prominence. Prominence is not assessed by stress or tonal contrasts, but by being longer and having more consonantal contrasts than prefixes (McDonough 2003: 7). If tonal contrasts were to be chosen as a property for defining prominence, both the stem and P-2 would receive the values ‘possible’, since these positions can exhibit all contrasts (high, low and falling tone). On the other hand, P0 (stem) has a lower property value; this shows that the anomaly of position P-2 can be justified by P-2 exhibiting properties that P0 does not have.

In Dumi, the morphs *k* (1pl) *n/Ø* (1sg > 2) and *ŋ* (1sg) (Van Driem 1993: 121) in Position 2 are consonantal. Their occurrence is lexically conditioned by conjugation class (pp. 132, 134). This position is not obligatorily filled since it exhibits only morphs involving first person. P2 morphs are a case of weak prominence status. Being consonantal, the stricter criterion applies whereby these morphs must be at a word edge to be considered prominent. This is not the case in Dumi, since other suffixes must always follow, indicating either number or tense (Van Driem 1993: 97-99; 106-117). Without this strict definition, these morphs would be one of the only suffixes that can fill the coda position of the syllables that include stems, which means that they are the only suffixes that are part of a stressed syllable. Furthermore, there is another LCI position in Dumi, the stem, which is always stressed – this position has a slightly lower property value, which means that the properties can justify the anomaly of P2.

In Supyire, the LCI morphs of P1, while never stressed, always appear at the end of the word. Furthermore, Supyire has another LCI position, the stem (P0), which is always stressed. The property value of P0 is also slightly lower than the one of P1, which may justify the explanation of the anomalies by the properties.

In Pilagá, P-1 has several other properties that can explain the anomaly. This position comprises two sets of indexation morphs which are part of every verbal construction. They can appear at the beginning of the word, if no other morphs (negation, object, indefinite subject) precede. P-1 is one of many LCI positions – P0, P2, P3, P4, all of which have the prominence value 'possible'. Furthermore, the two sets in Pilagá are mostly grammatically conditioned, and the fact that P-1 received the value LCI is due to the exceptional behavior in a few verbs. Vidal (2001: 137) shows that the LCI status of P-1 is weak, due to being almost semantically conditioned:

> "In the third group [of morphs from P-1], there is a fairly regular semantic contrast between the Set A marked and Set B marked verb forms (…). Though this schema accounts for the vast majority of the prefix choices on Pilagá verbs, the assignment of a particular case to a verb is still lexicalized."

Among the several other LCI positions, P-1 stands out as having the highest property value (3.5), which justifies the explanation of the anomaly by the properties.

In Nunggubuyu, morphs of P3 express tense, mood and indexation and are lexically conditioned. These morphs always appear at the end of the word since they occur in the last position. While these morphs never receive stress, stress in Nunggubuyu might not be a strong indicator of prominence. The stress rules by Heath (1984: 32) are summarized as follows:

> "Long vowels attract high intonation and stress, especially when adjoining syllables have short vowels." (…) "When the last few syllables of a word have short vowels, the penultimate attracts high pitch and a little stress, and a pattern of alternating high-pitched, stressed vowels in even-numbered syllables (right-to-left) may result."

The interplay of stress assignment may result in words that have several stressed syllables (7)

(7) Double stress in Nunggubuyu

ˈna:-ˈna-ni

1.exc.du>3pl-see-PT2

'We saw them.' (Heath 1984: 32)

Finally, Nunggubuyu has other LCI positions (P0, P1), which can be prominent. However, the property value of the anomalous position P3 is not higher than P0. Unlike P3, P0 is obligatory, which is understandable since this is the stem position.

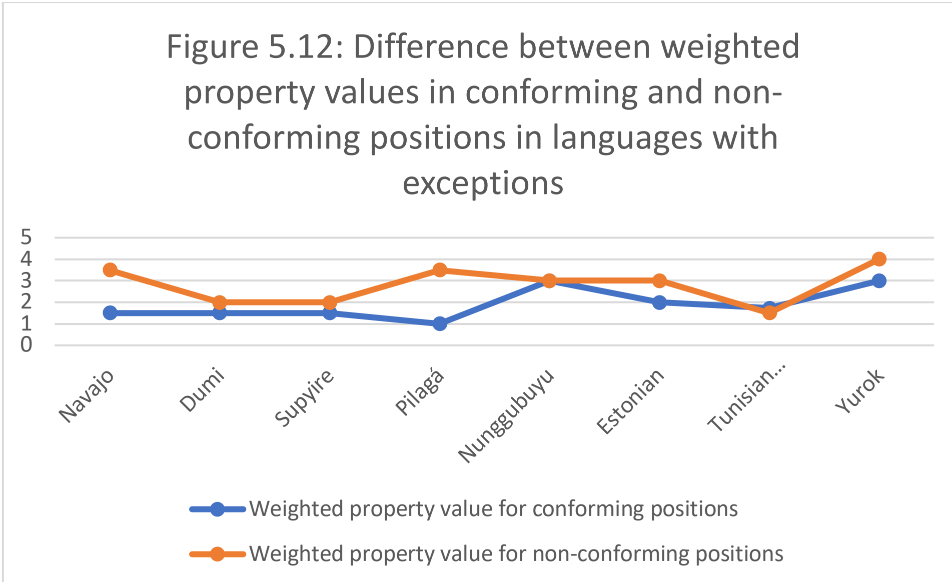
In Estonian, there are two morphological positions, and both can exhibit LCI. However, since stress falls on the first syllable, and P0 is always filled by the stem, P1 never receives stress. P1, as well as P0, always appear at word edges since no other positions precede or follow them. Nevertheless, P1 has a higher property value than P0, which justifies using these properties to explain the anomaly.

Present tense in Tunisian Zuwara Berber is in some conjugations realized by a vocalic suffix (*-a*/*-i*/*-u*) in P1, and it never receives stress. Besides occurring at the end of words if no other suffixes follow, it does not have other properties that could be considered prominent.

However, Zuwara Berber has a verbal template of 20 positions, and a record number of 9 LCI positions (including 4 LCI positions that are part of the stem), all of which can be stressed. However, the fact that P1 exhibits LCI but not stress can be regarded as an exception that does not defy the general tendency for LCI and prominence to co-occur in the entire template. Like in Nunggubuyu, P1 also does not have the highest property value because some stem LCI positions are obligatory, whereas P1 is not.

Finally, the LCI position P2.2 in Yurok received the prominence value ‘no’. This happened because these morphs only consist of either consonants or closed syllables with short vowels. According to Blevins (2003: 4), “[s]yllables with long vowels in Yurok always attract stress and are realized with a steady high pitch on the stressed syllable, while CVC syllables do not attract stress.” On the contrary, P2.1. can have morphs with long vowels, and therefore received the value ‘possible’. However, P2.2 exhibits several properties that can justify the anomaly. The indexation morphs in P2.2 always occur at the end of words, and P2.2 is always filled with morphs (obligatory). In fact, the morphs of P2.2 fill the same position (P2) as the morphs from P2.1, which are LCI morphs with the prominence value ‘possible’. According to Blevins (2003: 332-333), the morphs from P2.2 are not entirely unpredictable, but conditioned by some lexical formatives, which makes them almost phonologically conditioned and therefore weaken their status as LCI. P2.2 has a higher property value than P2.1, which justifies the explanation of the anomaly.

To justify the interpretability of these exceptions, a T-test analysis on the weighted values from Table 5.7 was conducted, both for the non-conforming (complex positions with the prominence value ‘no’) and the average value of the conforming positions of each language. This yielded the following graph (Figure 5.12).



The results are significant (p= .02889), which suggests that the anomalies can be justified by the presence of other properties pertaining to LCI positions in the languages. However, based on the qualitative criteria of the values, caution is advised when interpreting the results – the data itself may be overfitted to make the exceptions meaningful.

In conclusion, every deviation from the general correlation between LCI and prominence has been explained either by other prominent properties (word edges), properties that are susceptible for preservation of complexity (obligatoriness), by the fact that the exception does not go against the general tendency of a language to align LCI positions to prominent positions, or by the weak status of either LCI or prominence values. In most cases, the property value of the exceptions is higher than the average property value of the conforming LCI positions, which means that the properties can be used to explain the anomalies. The most defining trade-off of these exceptions is the presence of more than one LCI position associated with prominent syllables. That is, in every language where one LCI position is not prominent there is a LCI position that is prominent. In addition, the fact that many of these morphs occur at word-edges might explain why LCI is prosodically preserved these languages. In some cases, word edges might be a more distinctive prominent property than stress.

5.2.6.2 Always prominent UCI-positions

The other type of exception worth analyzing are UCI positions that have received the prominence value ‘yes’. There were only few instances in the LCI sample, namely P2 in Skou, P5 in Koasati and P4 in Betoi. Parallel to non-prominent LCI, some explanations are presented that can justify the universal tendency. However, prominent UCI is not as much of a surprise as is non-prominent LCI, since there are five instances of prominent UCI in the UCI sample. It appears that the absence of other inflectional categories increases the likelihood for UCI to attract stress, although there has been no instance where UCI received the value ‘yes’ in the GCI-sample. Within the LCI sample, there must be other reasons why UCI positions have received ‘yes’ values.

One reason could be that there are several positions that always attract stress, and words may have multiple prominent syllables (such as in Ket and Nunggubuyu). In these cases, having UCI that is always stressed might be the result of the likelihood for verbs to have an alternating stress pattern. Another factor for always prominent UCI positions might be that they are no longer affixes or clitics, but independent words which bear their own stress. Finally, UCI might be ‘weak’ UCI in the sense that it is almost NI – the semantics might be too specific. Morphs with specific semantics such as stems are more likely to attract stress, probably related to their propensity to be put into focus constructions (Boye and Harder 2012). Therefore, it is investigated whether one or more of these properties apply to the exceptions related to always-prominent UCI. Table 5.8 provides a summary.

	Template has multiple positions that can be prominent	**Independent word**	**Weak UCI status**
Skou P2	no	yes	no
Koasati P5	yes	no	yes
Betoi P4	yes	no	no

Table 5.8: Presence of further properties of UCI positions with the prominence value 'yes'. Due to the low number of these properties, and because the property 'template has multiple positions that can be prominent' applies to the entire template, the values of conforming UCI positions are not shown, and no unweighted total property value is indicated.

Table 5.8 shows that each position has at least one additional property that might explain the deviance from the general correlation. However, the scarcity of data makes it difficult to make general claims.

The negation morph *ka* in Skou is always stressed because it is analyzed as an independent word (Donohue 2004: 264). Donohue (2004: 82) writes that

> "stress is thus completely predictable, and is assigned to the first syllable in a simple word. In a word with proclitics, we find that stress remains on the first syllable of the root."

The morph *ka* has been annotated with the prominence value 'yes' since it is an independent word. However, Donohue (2004: 82) does not mention whether monosyllabic words have distinctive stress, indeed he writes that with regards to stress patterns "[t]he generally mono- or disyllabic nature of words in Skou also limits the amount that can be said." It could be that the word *ka* does not bear distinctive stress of its own, but following the description in Donohue (2004), it has been annotated as such. Donohue does not mention multiple stress placement in words, and the morph is unambiguously inflectional due to the value 'negative'. Furthermore, Donohue (2004) does not mention any conditioning on *ka*.

The Koasati modal morphs *–áhi* (intention) and *–á* (immediate intent) in P5 have been annotated as always prominent because they always carry a high tone. In Koasati, it is possible for a verb to have multiple prominent positions. In Example (8), the modal suffix *–áhi* receives a high tone, as does *há:lo* 'hear.'

(8) Existence of two high tones in one verb in Koasati

h**á:**lo-l-**á:**hi-k ca-bàn

hear-1sg-INT-SS 1sg.STAT-want

'I want to hear it.'; literally 'I want that I might hear it.' (Kimball 1991: 181)

While -*á:hi* and -*á* are not independent words, they have a weak inflectional status. First, 'intention' vs. 'immediate intent' are not common values for an inflectional paradigm. The reason why these morphs have been classified as inflectional is that they co-occur with other inflectional positions, such as tenses. Kimball (1991: 189) writes that

> "the suffix -á:hi- in combination with the tense suffixes produces compound tenses concerning actions that had the potential to be fulfilled, but were unrealized."

This brings up another property that shows the weak UCI status of these suffixes. *-á:hi* interacts with other inflectional morphs to form new grammatical categories. One could argue that *-á:hi* should be classified as GCI instead, but the meaning of *-á:hi* 'intention' is still evident in several constructions where it appears (wish constructions, polite questions, dubitative, potential to be fulfilled, pp. 180-183). Thus, the morph *-á:hi* does not change its own meaning in combination with other morphs yielding these functions. As such, it has been annotated as UCI, not as GCI. Nevertheless, this shows that *-á:hi* is not a prototypical example of UCI.

The last UCI position that was assigned the prominence value 'yes' is the negation suffix *-omé*/*-óme* in Betoi. It is always stressed, and forms part of verbs. Because verbs can be stressed in other positions, Betoi verbs can exhibit two stressed syllables, as is shown in (9).

(9) Existence of two stressed verbal syllables in Betoi

fala-ˈb-ome-ˈlu

say-3pl-NEG-1sg

'They are not speaking to me.'

There are no other properties that could explain why the exception is upheld by prominence. It must be said that Betoi is a scarcely documented language, and Zamponi (2003) does not list specific conditioning factors for the morph. The author describes the allomorphs as stressed on the last syllable but does not explicitly mention that they are always stressed. Because no other instances have been found of this suffix without stress, it has been annotated as UCI with the prominence value ‘yes’.

One could argue that prominent UCI is likely associated with multiple prominent syllables in words, or with independent words that have stress of their own. While these exceptions could be ‘explained away’ in each case, further conclusions cannot be made due to the scarcity of data.

5.2.7 Conclusion

Section 5.2 showed that there is a significant correlation between type of inflectional morphs and prominence. This holds also when excluding certain positions (stems) and investigating types of prominence. Across verbal constructions, more complex inflection tends to be expressed by more prominent units. Exceptions from the general pattern exist in some languages, but these exceptions have properties that might explain their deviation. Nevertheless, these exceptions might be just accepted as exceptions, since the research question does not posit an inherent functional relationship between prominence and conditioning – rather, complex morphology is more likely to be supported in prominent environments, and hypothesis H 1.1 does not predict that simplification of morphology *must* happen in less prominent environments. However, the correlation between prominence and morphological complexity shows that more fine-grained analyses of the interaction between morphology and prosody are fruitful. Section 5.3 investigates the relationship of complexity with the second independent variable, obligatoriness.

5.3 Complexity and obligatoriness

The second factor that has been associated with longer stability of linguistic structure is high type frequency, more precisely, deeply entrenched phonological and morphological schemas. As stated in Chapter III Section 3.4, high frequency of morpheme combinations increases ease of access (Bybee 2006). Evidence for the facilitative effect on children by high type frequency of morphs can be seen in proto-morphemic insertions (Peters 1997, 2001; Chee 2017). High type frequency of morphs not only leads to retention, but also to analogical extension, as Bybee (2006) shows with regularization of English irregulars (*wept* > *weeped*).

In addition, specific phonotactic patterns that are often repeated are also helpful in acquisition, as can be witnessed in children producing instances of phonological schemas (Vihman & Croft 2007: 708) and rhythm patterns (Grabe, Post and Watson 1999). Furthermore, highly frequent phonotactic schemas do not lead to erosion, like frequent individual phonemes, but are stable and stabilize morphology that aligns with the pattern. Frequent phonotactic patterns, such as open syllable restrictions in Central Pacific Oceanic (Blevins 2009), are strong schemas that prevent the change of syllable structure and might even initiate sound changes that align to this syllable structure.

In the methodology outlined in Chapter IV it was decided to focus on morphological paradigms that appear in every verbal construction. Obligatory paradigms strengthen both morphological and phonological schemas and are more likely to persist over time than morphs that do not belong to paradigms (Bisang 2014; see Chapter III Section 4.2.2). Thus, positions that exhibit obligatory paradigms have been annotated as obligatory, and positions that do not as non-obligatory. In order to account for strong phonological schemas, a distinction between obligatory positions whose morphs are always overt and obligatory positions with paradigms that exhibit zero morphs has been made. It is assumed that obligatory overt morphs in a certain position are deeply entrenched in the minds of speakers since they reinforce morphological and phonological schemas. Conversely, obligatory positions/paradigms with zero morphs are more entrenched on the morphological/semantic level, but less so on the phonological level since the positions are not always associated with phonological material, and as such they are not reinforced as phonological positions. Based on the assumption that obligatory positions reflect high type frequency and deeper entrenchment, and thus preserve linguistic structure, a general hypothesis is formulated as follows:

H 2.0: A higher complexity value is correlated with a higher rate of obligatoriness across verbal positions.

The values for complexity are again, LCI, GCI and UCI. The value for obligatory positions without zero morphs is 'yes', for obligatory positions with zero morphs is '(yes)', and the value for non-obligatory positions is 'no.'

5.3.1 General correlation

Analysis 2.0 – the first analysis on the relation between complexity and obligatoriness – has been run on the LCI sample, by including all types of positions (affixes and stems). A Chi-squared test was run to determine the significance. The results are presented in Figure 5.13.

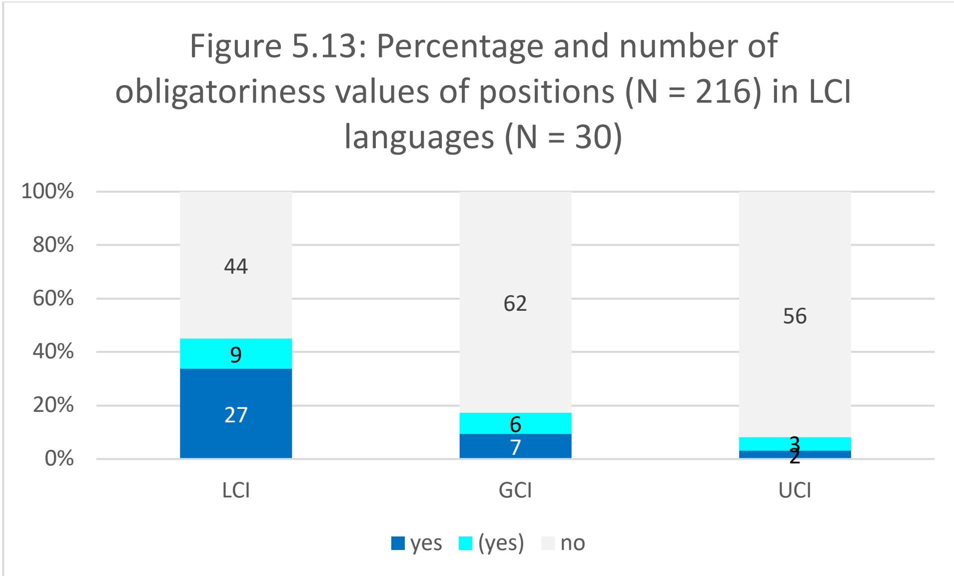


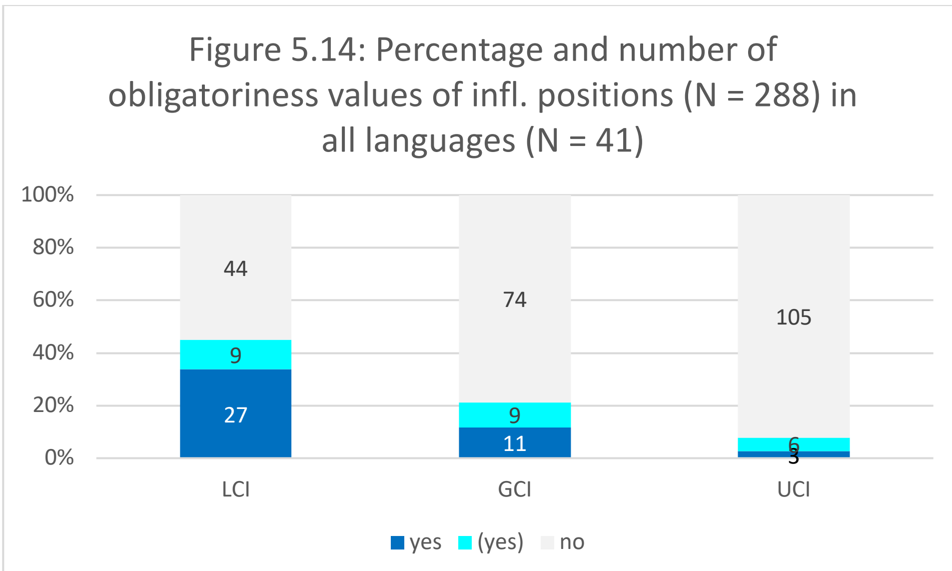
Figure 5.13 shows that there is a significant ($p < 0.00001$) association between morphological obligatoriness and inflectional complexity. LCI positions are more likely to be obligatory than GCI positions, both morphologically and phonologically, and GCI positions are more likely to be obligatory than UCI positions. While there is a positive association between obligatoriness and morphological complexity, the pattern differs from the pattern with prominence by exhibiting many more 'no' values. This suggests that obligatoriness across morphological positions is an exception, and that most inflectional positions, regardless of whether they exhibit paradigms, are non-obligatory. A language that shows this tendency in a nutshell is Ayutla Mixe (Romero-Mendez 2009); the verbal template is shown in Table 5.9.

Position	Categories and contexts of occurrence	Condition-ing	Obligatori-ness
-11	Subject/Object person markers. These in-dexation morphs are dependent on the A/M category of ‘dependence’. (pp. 294-295)	GCI	yes
-10	Non-nominal incorporation	NI	x
-9	Motion cum purpose	NI	x
-8	Directional and locative	NI	x
-7	Reflexive	NI	x
-6	Causative	NI	x
-5	Applicative	NI	x
-4	Benefactive	NI	x
-3	Incorporation	NI	x
-2	Part	NI	x
-1	Root (manner)	NI	x
0	Root, neutral completive, irrealis, independ-ent, dependent, completive. Different shapes for lexical conjugation classes (p. 338)	LCI	yes
1	Phase roots	NI	x
2	Desiderative	UCI	x
3	Inverse	NI	x
4	Perfective	UCI	x
5	Plural Allomorphs dependent on mood/as-pect (p. 319)	GCI	no
6	Independent, dependent, neutral, complet-ive, irrealis, cumulative morphs (p. 304)	GCI	yes

Table 5.9: Verbal positions, conditioning and obligatoriness of morphs in Ayutla Mixe (Mixe-Zoque; Robero-Mendez 2009).

Ayutla Mixe has three positions that are obligatory: Subject/object indexation in P-11; the stem position P0, and aspect/mood in P6 (Romero-Mendez 2009: 291). There are, however, six inflectional positions. Stems are the only LCI morphs in Ayutla Mixe. Grammatically conditioned morphs are present in P-11 and P6. Finally, there are two UCI positions: the desiderative morph *–ä'ä(n)* in P2 and the perfective morph *–n(e)* in P4. Thus, there are no obligatory UCI positions; two out of three GCI positions are obligatory, as is one LCI position. This proportion of alignment with obligatory positions reflects the global pattern, although the verbal template of Ayutla Mixe is unusual in having more than two obligatory positions.

To see the impact of the languages from the control samples, Analysis 2.0 has been run on all of the samples. The results for the GCI sample, which contains languages with only GCI and UCI positions, is shown in Figure 5.14.



Even across all samples, the trend behaves like in the LCI languages, where LCI are more likely associated with obligatoriness than GCI, and GCI more likely so than UCI. The results are again highly significant ($p < 0.00001$). Considering the behavior of obligatoriness across the samples, H 2.0, where more complex inflectional positions are more likely to be obligatory, can be confirmed.

Nevertheless, since stem positions are generally obligatory, if not by definition,[25] non-stem positions should be investigated separately. The reason is that stem positions, again, may exhibit other properties that favor morphological complexity, as pointed out in 5.2.2.

5.3.2 Non-stem positions

The fact that more complex inflection is found in stem positions may be an effect of obligatoriness, but it could also be the effect of the semantic or morphological prominence of stems. As such, Analysis 2.1 was conducted to show whether complex inflection is supported by obligatoriness outside of stem contexts. For this, a sub-hypothesis is formulated:

H 2.1: A higher complexity value is associated with obligatoriness across verbal non-stem positions.

Figure 5.15 presents the results obtained after excluding stem positions. Since stems are always obligatory, the results show a drop in the values 'yes/(yes)' in comparison to Figure 5.14. As such, the difference between obligatory LCI and GCI affixes is minimized for the LCI sample, since stem alternation is exclusively lexically conditioned.

---

[25] Non-obligatory stem positions might be smaller than the stem when they relate to certain segments that do not appear in certain contexts.

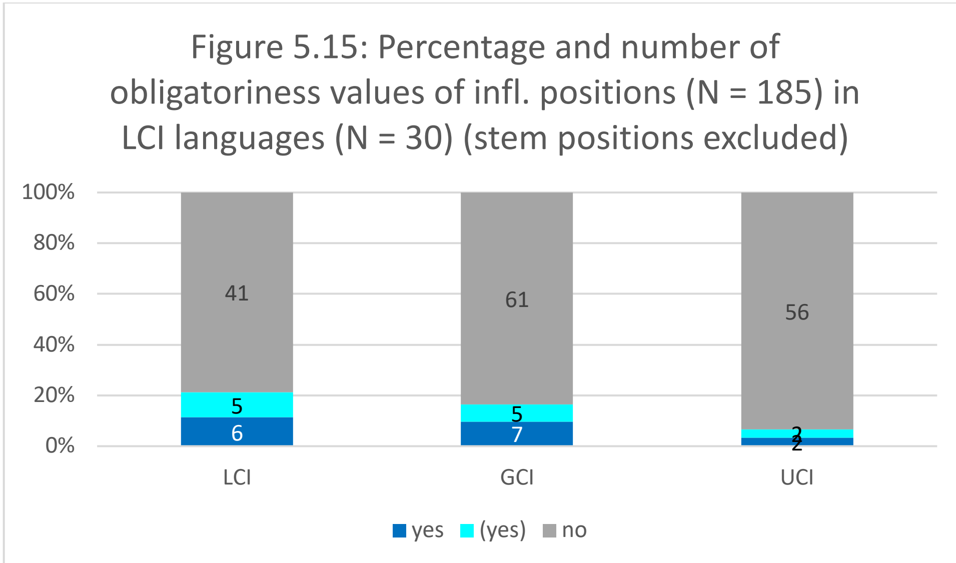
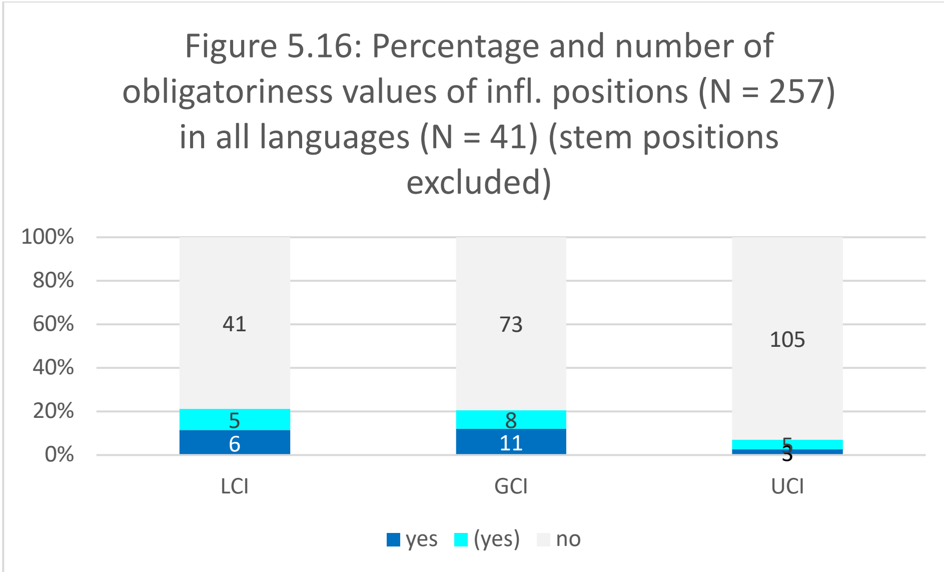


Figure 5.15: Percentage and number of obligatoriness values of infl. positions (N = 185) in LCI languages (N = 30) (stem positions excluded)

However, there is still a higher percentage of LCI positions (21.15 %) in comparison to GCI positions (16.44 %), and a higher percentage of GCI positions in comparison to UCI positions (6.67 %) that are obligatory. The results of the Chi-Squared test, when distinguishing all three possible values, are not significant (p = .27806). However, if the values 'yes' and '(yes)' are combined and contrasted with non-obligatory positions, the results approach significance (p = .081291). Thus, excluding stem positions from the calculation affects the pattern more than was the case with non-stem positions and prominence.

How does the picture look across all samples? Figure 5.16 shows the results when incorporating the 11 control languages that do not have LCI.



Including the control samples leads to an increase in obligatory positions for GCI from the 6 languages that have GCI but not LCI. This makes the percentage of obligatory positions ('yes' and '(yes)' values almost even for LCI (21.2 %) and GCI (20.7 %). Here, the results are slightly significant when distinguishing all obligatoriness values (p=.037469).

In conclusion, obligatoriness tends to be associated with higher complexity of conditioned inflection. H 2.1 can be confirmed both with regards to the pattern that includes and excludes stems. However, the fact that stem morphs are always obligatory reduces the number of tokens when comparing inflection across non-stem morphs. In non-stem positions, LCI and GCI categories behave alike, which means that the hypothesized association between higher complexity and obligatoriness only holds when regarding LCI and GCI as a unified category that is more complex than UCI. In general, one can say that more complex inflection is more likely obligatory, even if only the minority of positions receive 'yes' and '(yes)' values.

In the next section, the relationship between obligatoriness and fusion is closer examined, since this relationship might affect the association between complexity and obligatoriness.

5.3.3 Fused vs. non-fused positions

This section analyzes the relationship between obligatoriness and fusion – another aspect that might undermine the association between complexity and obligatoriness. Fused morphs are more complex than non-fused morphs, if they exhibit the semantic components of the morphs that were fused (cumulative morphs). Fused morphs account for many instances of GCI, as well as LCI expressed by stem alternation. However, the question arises whether the co-occurrence of obligatoriness and complexity is due to fusion *creating* obligatory inflection, rather than obligatoriness *preserving* fused morphs. For example, it might be the case that fusion of two non-obligatory paradigms results in an obligatory paradigm. In some languages, alternative analyses of obligatory paradigms render several non-obligatory paradigms. The following example is given from morphs of P-2 in Navajo. According to Young and Morgan and Midgette (1992), there are two analyses for representing morphs in this position. These two analyses are shown in Table 5.10.

	Zero-imperfective		Ni-imperfective		Yi-imperfective (progressive)	
Subject features	surface	underlying	surface	underlying	surface	underlying
1sg	yish-	Ø-sh-	nish-	ni-sh-	yish-	yi-sh
2sg	ni-	Ø-ni-	ní-	ni-ni-	yí-	yi-ni-
3	yi-	Ø-Ø-	yí-	ni-Ø-	yi-	yi-Ø-
1du/pl	woh-	Ø-oh-	noh-	ni-oh-	wooh-	yi-oh
2du/pl	yiid-	Ø-ii-	nii-	ni-ii-	yii-	yi-ii-

Table 5.10: Surface and underlying manifestations of Navajo P-2 morphs.

Table 5.10 shows three imperfective paradigms. Following the surface-form analysis, these paradigms are obligatory since there is always phonological material present in P-2. The more structuralist analysis posits two positions for P-2, with both of them potentially exhibiting zero morphs. In this dissertation, only the surface form has been considered, which results in a single TAMI position, namely P-2. P-2 is therefore obligatory since it always exhibits overt morphs. Intra- and interlinguistic evidence show that historically, P-2 paradigms have emerged from two paradigms belonging to aspect and person, since they are presented as different affixes in reconstructions of earlier stages of Athabascan (Vajda 2019; Kari 1989; Krauss 1965). This shows that obligatoriness can be the result of fusion rather than a force impacting the longevity of fusion.

In order to grasp the effect that obligatoriness has on the retention of morphs, one would have to compare the rate at which paradigms fuse and their relative longevity across languages.

If paradigms that become obligatory by fusion remain obligatory *and* fusional (i.e., morphologically complex) over a long period of time, the effect of obligatoriness on morphological complexity could be proven. If paradigms which have become obligatory by fusion do not remain obligatory or complex for a long period, the effect of obligatoriness on morphological complexity cannot be proven. Obviously, this represents a separate study and will have to involve different variables, such as rate of change, rate of loss of obligatoriness and complexity, etc. What can be analyzed in this study, however, is whether obligatoriness is distributed across fusional and non-fusional inflectional positions.

It could be argued that if the distribution of obligatoriness was equal across conditioning types regardless of fusion, fusion would not be a significant factor in the relationship between morphological complexity and obligatoriness. However, if obligatoriness is indeed correlated with fusion, this would not disprove the preserving effect of obligatoriness on complexity. In the following, fused GCI and LCI positions are compared with non-fused GCI and LCI positions in relation to obligatoriness. Positions were considered whose morphs manifest a fusion between the conditioning and conditioned element, that is, the elements that resulted in assigning complexity values to the positions. Thus, it has been investigated whether the LCI morphs derive from lexical elements (stem, thematic morphs) and inflectional elements, and whether GCI morphs derive from morphs that are associated with the conditioning and the conditioned grammatical category.

One problem with detecting fusion is that one cannot know whether one morph historically emerged from two morphs. Not all morphs that contain different semantic features (cumulation) derive from morphological fusion. Indexation morphs almost always cumulate number and person, but they do not always derive from morphs fusing together that expressed person and number, at least not within recoverable history. For example, Heine and Song (2011: 612–617) show that many first-person plural pronouns derive from nouns meaning ‘people’. Likewise, not all morphs that emerged from fusion are cumulative. For example, the French interrogative particle [kɛsk] derived from the fusion of the words of the phrase ‘que + est + ce + que’ (‘what is it that’) (Foulet 1921: 266-267). The morpheme [kɛsk] no longer distinguishes these meanings.

In order to investigate the relationship between obligatoriness and fusion among LCI and GCI positions, I have followed a naïve approach: those morphs for which authors provide further formal segmentation or historical reconstructions that indicate that the conditioning and conditioned features of the morph were overtly expressed at an earlier stage will be annotated as ‘fused’, those morphs for which authors do not provide evidence for a further analysis will not be annotated as fused. Furthermore, I have regarded those morphs as fused whose formal make-up strongly indicates a combination of conditioning and conditioned elements, as established by comparison to other paradigms. For example, Azari has two negative morphs, *-mA* (Negative) and *-(y)AmmA* (Impossibilitive). While the author does not explicitly mention it, the form *-(y)AmmA* bears formal resemblance to modal morphs *-(y)A* ‘optative’ *-(y)Abil* (ability) *-(y)AcAX* (intentional, future). Also given the semantic resemblance (impossibility can be decomposed as possibility and negation), the morph *-(y)AmmA* has been annotated as a fused form. In order to obtain meaningful results, the following exclusions have been made in the comparison:

- No UCI positions. UCI positions are rarely obligatory and almost never derived from fusion (except indexation morphs co-expressing person, number, gender or clusivity that have been included in this category according to Chapter IV

Section 4.5.3.1). If UCI morphs were to be derived from fusion, they would not inherit cumulation of conditioning and conditioned features.

- No stem positions. Inflectional stem morphs are always obligatory because they are the position that is always filled. A stem morph does not become obligatory because it fuses with other morphs, but because it is the lexical requirement of the verbal construction. Thus, neither the question of whether fusion leads to obligatoriness nor whether obligatoriness preserves fusion (and as such, morphological complexity) can be answered by including stem morphs.

By excluding these types of positions, it was expected that the relationship between obligatoriness and fusion could be established on a neutral basis. In total, 143 positions, all of which are LCI and GCI non-stem positions, were compared to find evidence for fusion between the conditioning and conditioned categories. The hypothesis pertaining to this analysis is formulated as follows:

H 2.2. Obligatory positions are more likely to show evidence of fusion of more than one morpheme.

The results pertaining to Analysis 2.2 are presented in Figure 5.17.

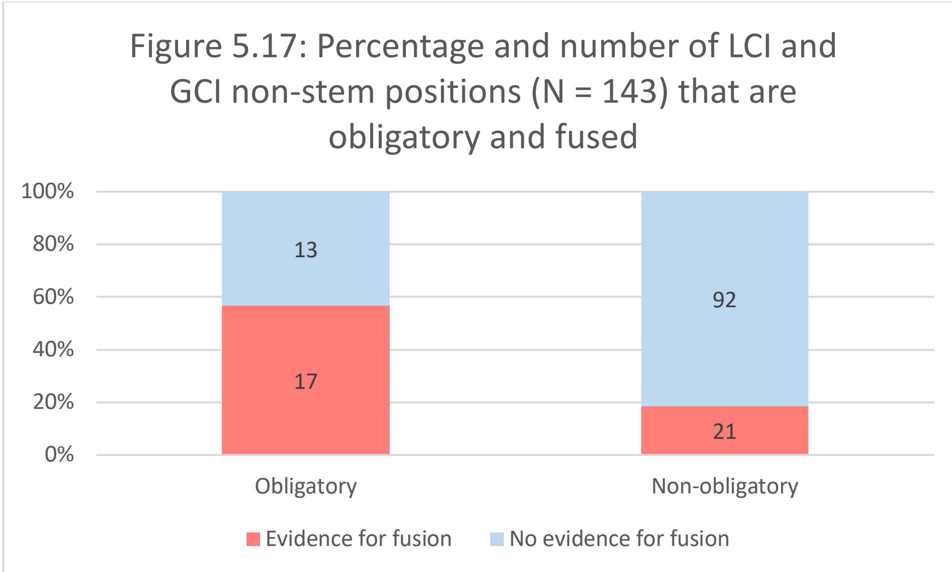
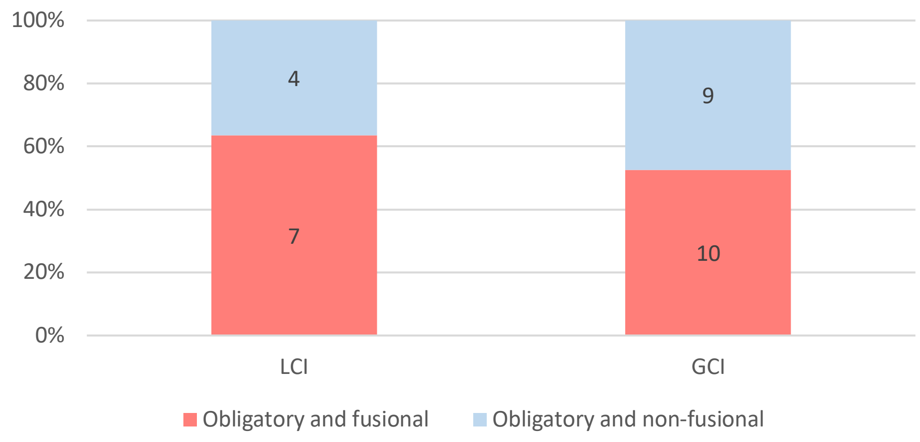


Figure 5.17 shows that H 2.2 can be confirmed: Obligatoriness of complex positions (LCI and GCI) is associated with evidence for fusion of conditioning and conditioned categories in these positions. The Chi-Squared test is significant (p = .000027), although the result says nothing about directionality. Figure 5.17 does not show which categories involved in the fusion rendered a morph obligatory. In fact, there are 21 instances of fusional morphs that are not obligatory, and 13 out of 30 obligatory morphs are non-fusional. While obligatoriness and evidence for fusion have been shown to occur together, this does not mean that fusion is the precondition of obligatoriness. This relation does also not rule out an effect of obligatoriness on fusion *after* fusion happened. A closer look at the distribution of fusional positions among obligatory LCI and GCI positions reveals that there is not much difference between LCI and GCI (Figure 5.18).

Figure 5.18: Percentage and number of fusional positions among obligatory LCI and GCI positions



A slightly higher proportion of LCI positions are associated with fusion than of GCI; however, this trend is not statistically significant (p = .557765), possibly because of limited data. This means that there is no certainty that fusion is a factor that contributes to the higher likelihood for more complex inflection like LCI to be obligatory in contrast to GCI. In conclusion, obligatoriness is associated with fusion – even for non-stem morphs. This does, however, not confirm that obligatoriness is a weaker factor than fusion in contributing to the distribution of morphological complexity across verbal templates.

5.3.4 Conclusion

More complex inflection is more likely expressed in obligatory positions, which confirms H 2.0. However, several factors contribute to this association. On the one hand, stems are always obligatory, and exhibit also other traits that can be associated with longer retention of morphological complexity, such as semantic and morphological prominence. Nevertheless, the association between obligatoriness and complexity across all samples is still significant when excluding stem positions. On the other hand, obligatoriness is often associated with formal fusion of positions, which means that in some cases, obligatoriness might not be a preserving factor for morphological complexity, but rather the result of morphological complexity (in terms of fusion). However, the proportions of fused and obligatory positions do not significantly differ among LCI and GCI positions. This also does not minimize the effect that obligatoriness might have on the perpetuation of LCI and GCI morphs. Unlike obligatoriness, fusion alone may not have a stabilizing potential. Fusion is more likely the result of than the cause of high frequency.

5.4 Prosodic prominence and obligatoriness – independent variables?

Finally, a closer look at the interplay between the two structure-stabilizing parameters – obligatoriness and prominence – should be taken. The results obtained in the previous sections show that prominent syllables and obligatory positions are associated with the distribution of inflectional complexity, pointing to the impact of these structural units on the stability of complex morphology. However, proving that the effect of prominence and obligatoriness are independent from one another could shed light on the existence of the general stabilizing potential of various structural units; one could say that prominent syllables are as much a factor as obligatoriness, and maybe, there are further factors. Thus, the correlation between more complex positions and prominence/obligatoriness might be due to a correlation between the stabilizing variables themselves. On the other hand, a positive correlation (more prominent positions are also more likely to be obligatory) could either mean that morphological complexity is more likely maintained when two variables interact, or that a third factor motivates the correlation between these two structural variables. The two variables should be compared with regards to the general pattern, and with regards to specific types of positions (stem vs. non-stem; complex vs. non-complex).

Analysis 3.0 has been run on values of all the samples together, to test the association between prominence and obligatoriness across all morphological positions. This includes non-inflectional positions (NI) which were also coded for the values of the independent variables but were not investigated in the earlier analyses. While any type of non-inflectional positions can be associated with prominence, the only NI positions that were obligatory were stems. In order to simplify the analysis, and because the 'no' values of obligatoriness prevail greatly, obligatory values 'yes' and '(yes)' have been merged to 'yes'. The results of the Analysis 3.0 are presented in Figure 5.19.

Figure 5.19: Obligatoriness and Prominence in positions (N = 474) across all languages (N = 41)

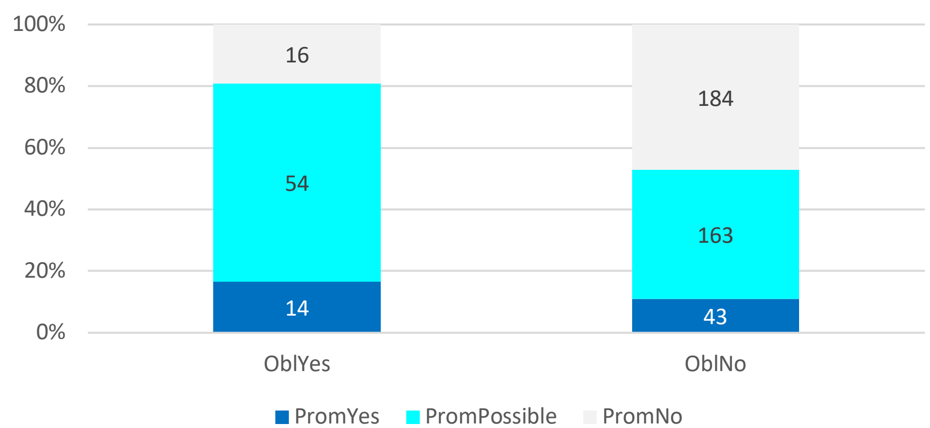
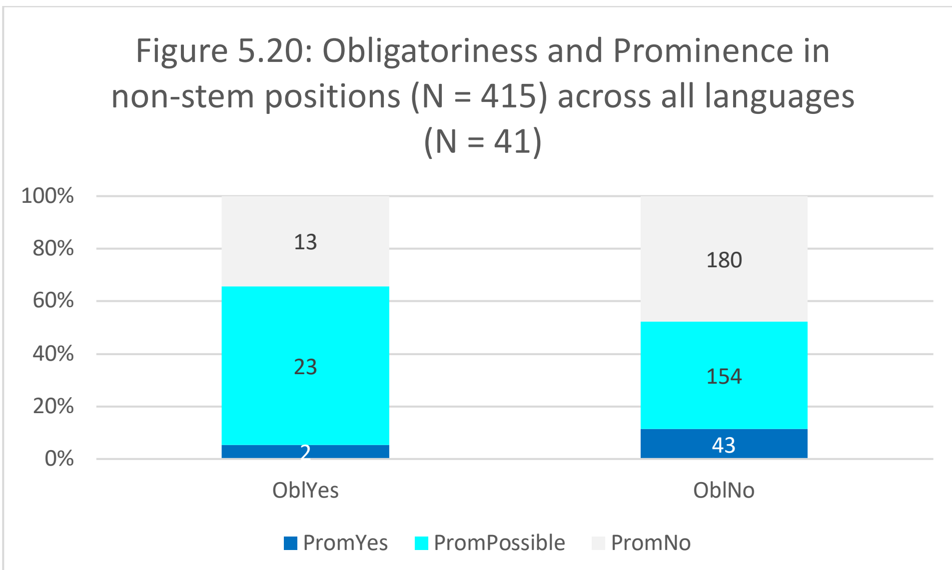


Figure 5.19 shows that across all samples, obligatory positions are proportionally more likely to be prominent. The difference is highly significant (p = .000013). This shows that obligatoriness and prominence as defined in this thesis might not be independent variables. One could

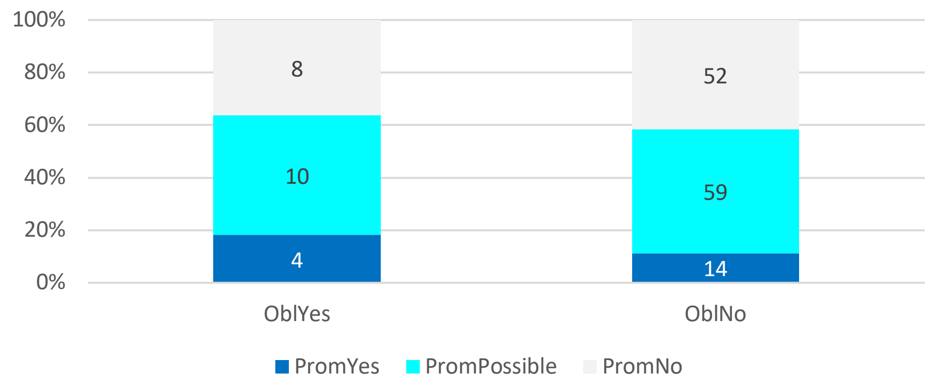
conclude that obligatory positions more likely attract prominence than non-obligatory positions. The reverse scenario where prominence attracts obligatory positions is not plausible. However, considering the specific status that stems have in relation with prominence, and more so with obligatoriness, it could be contemplated that the reason why obligatoriness and prominence co-occur is that this co-occurrence is mostly confined to stems. Therefore, Analysis 3.0 has been run on non-stem positions. The results are presented in Figure 5.20.



The results presented in Figure 5.20 suggest that obligatoriness is associated with prominence. Obligatory non-stem positions are still more likely to be prominent than non-obligatory positions, but this difference only trends towards significance (p = .05799). Stems are, as expected, a relevant factor that contributes to the association between obligatoriness and prominence, but one cannot say that these variables are dependent in non-stem contexts.

These results point towards the following conclusion: prominence and obligatoriness each provide structural effects for the distribution of complex morphology, but these effects are stronger when the two properties combine. Furthermore, stems are a third factor that might explain why obligatoriness and prominence are correlated; however, since stems are also considered stable morphological environments, this convergence can be interpreted as stems additionally strengthening the environment to be stable and thus promoting the survival of complex morphs that are expressed by stem alternation. Also, one could assume that there is a close association between prominence and obligatoriness because of the nature of the samples: 30 out of 41 languages exhibit LCI. Possibly prominence and obligatoriness map onto one another because LCI is more likely to occur in environments that comprise both of these structural properties. This does not mean that LCI attracts obligatoriness and prominence, it means that the co-occurrence of these properties is the most favorable structural environment for the retention of unpredictable, non-generalizable inflection. Because of this possibility, Analysis 3.0 was conducted only on the languages from the GCI and UCI sample. While GCI languages still exhibit complex inflection, they will be included with UCI languages in this analysis, in order to see the effect of the exclusion of languages with the highest type of complexity. Figure 5.21 presents the results.

Figure 5.21: Obligatoriness and Prominence in non-stem positions (N = 147) in GCI and UCI languages (N = 11)



The graph suggests that there is no distinctive association between obligatoriness and prominence in GCI and UCI languages. The statistics corroborate this observation since the Chi-Squared test is not significant (p =.643218). This means that in languages that do not have LCI, there is no association between obligatoriness and prominence, and the variables can be considered independent. One can interpret these variables as independent variables that happen to converge in LCI languages. However, this convergence may not be caused by a language having LCI; rather, the survival of LCI is facilitated by this convergence. Furthermore, the convergence between prominent, obligatory and stem positions favor the maintenance of complex morphology even more. In Chapter VII, the additive effect of structural stabilizing units will be further theorized.

5.5 Conclusion

Three analyses including sub-analyses concerning the structural interplay between complexity of conditioned inflection, prosodic prominence and obligatoriness have been conducted. The results presented in this chapter show significant associations between prominent properties of verbal positions and morphological complexity, as well as obligatoriness of verbal positions and morphological complexity. The overall picture is that prominence and obligatoriness co-occur with higher complexity of inflectional morphology.

Higher complexity is more likely to be expressed by prominent syllables/consonants; this holds for the LCI and GCI sample, which shows that LCI is the type of inflection that is more often prominent, but less so for GCI, and even less so for UCI. This general trend also holds after excluding stem morphs, which are likely to attract prominence. The trend is not contradicted by comparing different types of prominence assigned to positions, such as stressed positions, positions with high tonal variability, and positions with heavy syllables; nor is it contradicted by the behavior of positions with consonantal morphs. All these types of prominence are more likely to be distributed among LCI and GCI than among UCI positions. Furthermore, the number of complex inflections (LCI, GCI) is related to the number of possible prominent positions across the languages. This means that a language where many morphological positions can be prominent has a higher chance to keep complexity in words. Finally, outliers from the main trend (never prominent LCI; always prominent UCI) can be explained by them having other properties that explain the anomalies, unlike positions that conform to the trend.

The association between complexity of inflection and obligatoriness also confirms the hypotheses. More complex positions (LCI, GCI) are more likely to be obligatory than UCI. However, this tendency is not as strong as the association with prominence; several morphological positions had to be excluded (stems, non-inflectional positions) to see the effect of obligatoriness on complexity. This effect is not significant when excluding stems from the LCI sample, and only slightly significant in the overall sample, probably because of the lack of ‘yes’ and ‘(yes)’ values for obligatoriness. This lack means that a distinction between ‘yes’ (position is always filled with morphs) and ‘(yes)’ (position has an obligatory paradigm, but with zero morphs) values is not really useful. Furthermore, obligatoriness might itself not be a driving factor behind the distribution of morphological complexity. Instead, it could be the effect of complexity itself, that is, obligatory positions might result from paradigms undergoing fusion. However, this does not prove that obligatoriness is *not* a driving force behind the retention of complexity. It shows that general type frequency is a more likely factor than obligatoriness to explain the retention of morphology, since obligatory paradigms, whether fused or not, are an indicator for high frequency.

Given the positive association between prominence and obligatoriness, one might conclude that the variables prominence and obligatoriness are not independent from one another. However, it could be shown that stems and LCI languages themselves are responsible for this positive association, and that in languages with less complex morphology (GCI, UCI), there is no significant relationship between prominence and obligatoriness. This points to another trend in morphological complexity. The more complex morphology a language has, the more facilitating and stabilizing properties (stems, prominence, obligatoriness) converge in one position. Complexity of inflection is, according to this idea, an epiphenomenon that is propagated through facilitative and stabilizing contexts, and the more these properties are comprised in this context, the more likely morphological complexity prevails. Future studies might help to detect and operationalize additional structural properties in language that favor certain structures. The takeaway from this chapter is that morphological complexity might not be beneficial for speakers per se, but structural contexts associated with facilitation provide many hooks to acquire and retain it. This insight allows for a further theoretical examination on the relationship of inter-structural relationship in language, which will be provided in Chapter VI. A concluding discussion on the implications of this study will be presented in Chapter VII.

CHAPTER VI: TOWARDS AN EVOLUTIONARY THEORY OF STABILIZING STRUCTURE

6.1 Introduction

Chapter V has shown that complex inflection is associated with stabilizing structures in language, suggesting that the distribution and perpetuation of complex morphology is not due to something inherently stable about conditioned morphs, but due to effects associated with structure that interact with these morphs (prominent and frequent environments). This chapter reflects on this structural interrelationship between morphology, prominence and frequency, and proposes a theory that is embedded within a usage-based, evolutionary framework. It is necessary to provide a theoretical interpretation of the results since structural interaction is usually analyzed by formal approaches. In the functionalist literature, 'external' ecological factors such as social, cultural or natural factors are often invoked to explain structural variety and change (see Mufwene 2018; De Busser 2015), whereas explanations that consider how structure itself is a factor (the 'internal ecology' of language) has received comparatively little focused attention. In the following sections, it is argued that a complex adaptive, evolutionary perspective of language allows to address structural interaction as a process resulting from replication and selection of linguistic material. Croft's (2000) Theory of Utterance Selection provides a framework that situates language structure as the result of speakers interacting with the environment. What is replicated is phonological and morphosyntactic structure pertaining to utterance, and these linguistic replicators are termed 'linguemes' by Croft (2000: 28). What this dissertation seeks to interpret is the interaction between structural entities, that is, how the replication between different types of linguemes influence one another.

After proposing different perspectives on how these structural effects are related to the processes of interaction, selection and replication, it is concluded that the most plausible interpretation is the one where replicated structure, such as prominent and frequent linguemes, are also part of the environment with which the speaker interacts; this interaction causes differential replication of other linguemes in utterances.

6.2 An evolutionary approach to utterance selection

While this dissertation has researched synchronic distributional properties between different structural elements, the theoretical implications are diachronic: the concept of stability invokes dynamicity, and therefore the influence of prominence and frequency on complexity must be interpreted as a dynamic relationship unfolded in successive usage events. Many usage-based linguists view language as a 'complex adaptive system,' a system that is not stable, but constantly adapting and dynamic (Bülow, de Bot & Hilton 2017; Beckner et al. 2009; Van Geert and Verspoor 2015; Mufwene 2008, Steels 2000). In their position paper, Beckner et al. (2009) define the key features of this language system:

> "(a) The system consists of multiple agents (the speakers in the speech community) interacting with one another. (b) The system is adaptive; that is, speakers' behavior is based on their past interactions, and current and past interactions together feed forward into future behavior. (c) A speaker's behavior is the consequence of competing factors ranging from perceptual mechanics to social motivations. (d) The structures of language emerge from interrelated patterns of experience, social interaction, and cognitive processes." (Beckner et al. 2009: 2).

As (d) suggests, this framework allows the investigation of all sorts of interactions. However, it is not really clear whether the structural interaction shown in this dissertation is included, since language structure is seen as a product of several interactions, and not as an interacting agent. If structural elements such as stressed syllables and frequent morphological templates are responsible for the stability and change of other structural elements (complex morphs), a complex-adaptive view of language would need to regard structural interaction as a type of 'patterns of experience, social interaction, and cognitive processes.' In order to approach the structural effects from the perspective of a complex adaptive system, I have decided to pursue an analysis based on Croft's (2000) Theory of Utterance Selection, in which the role of language structure is clearly defined. According to this theory, linguistic structure emerges and changes from the selection and propagation of specific utterances by human beings as a consequence of the interaction of speakers and the environment. The Theory of Utterance Selection is a manifestation of the neo-Darwinian General Analysis of Selection proposed by Hull (1988).

6.2.1 Hull's General Analysis of Selection

The General Analysis of Selection (GAS) was formulated by Hull (1988) as a generalization of ideas of selection in biological evolution. For example, GAS can explain how scientific ideas evolve. Evolutionary change in the GAS is change by *replication*, which involves *replicators* and *interactors*, creating *lineages*. The entities partaking in replication are presented below, with the definitions given by Hull (1988: 408-409)

(1) 'REPLICATOR – an entity that passes on its structure largely intact in successive replications.'

(2) 'INTERACTOR – an entity that interacts as a cohesive whole with its environment in such a way that this interaction *causes* replication to be differential.'

(3) 'LINEAGE – an entity that persists indefinitely through time either in the same or an altered state as result of replication.'

Hull (1998: 217, 412-418) describes, citing Mayr (1978: 52), evolution governed by two causal mechanisms: *generation of variation* and *selection*. Generation of variation is the result of *altered* replication of replicators and selection involves *differential* replication. Altered replication happens when the replication of one replicator leads to new structural variants of that replicator. An example from biology is when a gene (*replicator*) is copied (*replicated*) but this copying process leads to differences from the original copy (*altered replication*). On the other hand, differential replication happens as a result of selection between multiple competing replicators, where one replicator is selected over others based on reasons that lie in the interaction of the interactor with their *environment*. Hull (1988) does not provide a distinct definition for the *environment* as the fourth entity in the process, despite playing a role of equivalent significance in interaction alongside the interactor. This might be due to the environment being composed of several elements or parts, which are variable (p. 467).

In biology, differential replication happens when gene X is selected over gene Y and becomes predominant in the gene pool. As a result, there are more replications of gene X than of gene Y (*differential replication*). The reasons for why there is the change where gene X is replicated differently to gene Y cannot be explained by 'copying errors' but by factors that lie in the interaction between the organism (*interactor*) and its ecology (environment, which includes other organisms), leading to pressures that determine the fitness of the genes and the favoring of one gene over others (*selection*).

Hull (1988) argues that the scientific process is itself a manifestation of the GAS; the replicators are the scientific concepts that get passed on. Hull (1988: 434) argues that scientists "are the primary interactors in the conceptual development of science" and the replicators are "elements of the substantive content of science–beliefs about the goals of science, proper ways to go about realizing these goals, problems and their possible solutions, modes of representation, accumulated data, and so on."

Having laid out how evolutionary selection can be generalized to other disciplines, the next section will explain how the GAS applies to language, an enterprise pursued by Croft (2000).

6.2.2 Croft's Theory of Utterance Selection

Croft (2000, 2013) formulates the manifestation of the GAS in language as the Theory of Utterance Selection (TUS). This evolutionary approach assigns the social-communicative context the role of the environment, the speaker the role of interactor and the *lingueme* the role of replicator. Croft (2000) uses the term 'lingueme' for "the paradigm linguistic replicator in language (…) [.] An utterance is made up of linguemes, and linguemes possess linguistic structure." (p. 28). A lingueme refers to phonological or morphosyntactic structure of utterances as seen from its evolutionary perspective, as the result and the potential of being replicated. Linguemes are spatio-temporally bounded entities that are replicated in utterance events. These linguemes form lineages across the lifetimes of speakers and across generation of speakers.

Table 6.1 shows further instantiations of the GAS in biology and language (Croft 2000: 38), with the entities that are relevant for this Chapter in bold.

Generalized theory of selection	*Paradigm instantiation of selection in biology*	*Paradigm instantiation of selection in language*
replicator	**gene**	**lingueme**
replicators in a population	gene pool	lingueme pool
structured set of replicators	string of DNA	utterance
normal replication	**reproduction by e.g., interbreeding**	**utterance production in communication**
altered replication	**recombination, mutation of genes**	**mechanisms for innovation**
alternative replicators	alleles	variants
locus for alternative replication	gene locus	linguistic variable
interactor	**organism**	**speaker (including grammar)**
environment	**ecological environment**	**social-communicative context**
selection	**survival and reproduction of organisms**	**entrenchment of convention by speakers and its propagation in communication**

Table 6.1: Instantiations of the General Analysis of Selection in biology and language

Altered replication (generation of variation) and differential replication (selection) both occur in the evolution of language. When speakers produce utterances, the linguemes can be replicated unfaithfully with regards to their previous variants, and this creates new variants. To illustrate this, we can refer to the different lineages showing the change from *vuestra merced* towards *Usted* in Figure 1.2 (Chapter I). This change involved the existence of multiple variants such as [vwes], [vus], [vwes], [voas], [us], and many more. The reason why these variants were created is due to altered replication.

Nevertheless, if altered replication was the only mechanism of change, we would see myriads of variants co-exist at any point in time. In fact, the utterance [usteð] is the most prevalent variant that has survived from the pool of multiple variants, and this is due to selection, which involves *differential* replication. As mentioned above, selection manifests by the preference of one variant over another, and this bias leads to a change in the distribution of utterances and the 'lingueme pool', creating lingueme lineages of different length and frequency. Croft (2000: 32) argues "that the selection process is essentially a social one, and not a functional one in the sense of (external) function". Nevertheless, the reasons for the social processes leading to selection of one variant can be influenced by functional-adaptive properties that facilitate communication. Thus, both functional and social factors influence the emergence of linguemes and their lineages.

Figure 6.1 visualizes the relationship between the environment, speakers, utterances and linguemes, and their respective processes they partake in (interaction, selection, altered and differential replication).

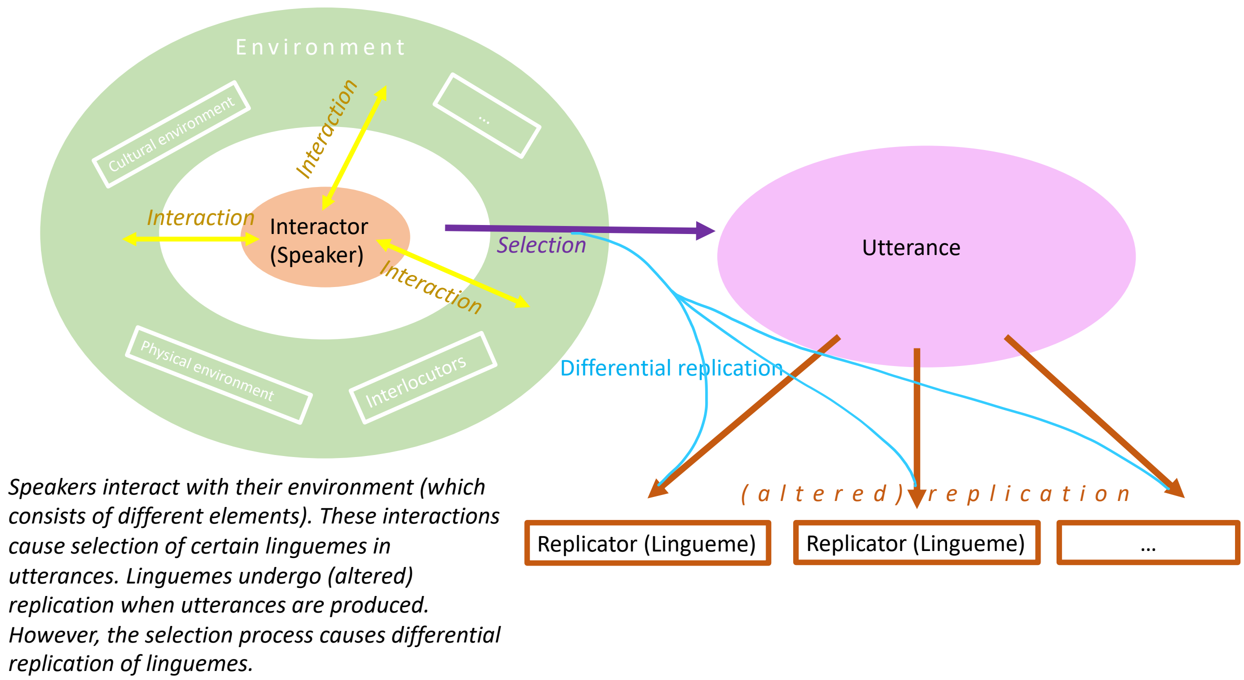


Figure 6.1: The Theory of Utterance Selection visualized

While altered replication happens "autonomously", i.e., linguemes change without any other lingueme competing, the changes in those variants can be explained by functional constraints. For example, the altered replication by which case suffixes in English eroded can be explained by the constraint of efficiency: speakers reduce the length of words and morphs when they are highly frequent and understood from context. Nevertheless, the result of altered replication – having multiple variants –, is not functional. On the other hand, the result of selection – one variant predominating another in speech – can be interpreted from a functional perspective as it shows the preference of speakers based on facilitating communication. Furthermore, selection processes are reflected in the distribution of linguemes in different environments.

It is the idea in this dissertation that the 'differential distribution' of inflection (the tendency of LCI and GCI to be found in prominent and frequent environments) can be explained as a result of the differential replication of linguemes in environments that are more likely preferred by speakers. Thus, I argue that structural effects are not only the product of altered replication of linguemes, but that they are connected to selection processes in speakers that can be both socially and functionally motivated, and both influence the structural stability.

Some phenomena can be directly explained by functional or social principles. For example, one possible explanation for the devoicing of German codas is the functional principle that suggests it is challenging to pronounce voiced obstruents in coda position, as stated by Haspelmath (1999: 194) and Keating, Linker & Huffman (1983). This suggests a specific interaction of speakers and environment (in this case, constraints on articulation pertaining to the bodily environment of speakers). Thus, one could imagine a scenario in which speakers have the choice between producing an utterance with one of two lingueme variants, one with voiced, another with voiceless obstruent codas, and the latter is chosen. However, this scenario no longer exists in Standard German. More likely, the continued usage of voiceless obstruents and the lack of reemergence of voiced obstruents is attributed to adherence to convention, another crucial aspect of the interaction between speakers and their environment. This is a social principle where the particular convention is valued in a specific speech community, and as such preferred over variants that are generated due to altered replication. In both cases, although the change from voiced to voiceless segments can be interpreted as the result of altered replication, the reasons for its maintenance ('normal' replication) can be seen as the result of selection governed by functional or social principles.

Nevertheless, environment-based explanations might be harder to achieve when dealing with less functional variants, especially if they remain in language for a long time. As Croft (1999: 208) notes in a reply to Haspelmath (1999), "a model in which selection is functionally ‘adaptive’ cannot account for the selection of ‘dysfunctional’ variants.” This leaves less functional variants to be explained by purely social factors and would mean that they would be found in cross-linguistically less predictable environments, because social conventions are not universal, and uncorrelated to functional principles. The distribution and stability of complex conditioning, which can be viewed as ‘dysfunctional’, would thus be merely the product of speakers abiding to social conformity. This conclusion is however unsatisfactory. It does not explain why certain difficult-to-learn variants are rare and others more frequent. If social convention were the only factor for the stability of complex morphology, we would expect it to be randomly distributed in utterances, and its lineages to be of variable length. This would mean that the distribution would be merely governed by the process of replication, in this case, normal and not altered replication. Instead, the empirical part of this dissertation has shown that the association between morph conditioning types, reflecting complexity, and other linguemes such as prominent syllables and entrenched morphological templates, is meaningful. Thus, social factors like conformity to convention cannot fully account for the existence and perpetuation of complex morphology, and neither can functional factors (if variants are indeed dysfunctional).

Chapter 6.3 has suggested, based on the literature, that the distribution of complexity in utterances can be explained by certain parts of utterances being more prominent and entrenched in speech. These properties are functional-adaptive: entrenchment benefits speakers by facilitating production and prominence benefits hearers by facilitating perception. However, it has been argued that prominence and entrenchment are not alone responsible for this stability. Rather, the stabilizing effect of prominence and entrenchment is dependent on certain *structural entities* being prominent and entrenched, such as syllables, word edges, or morphological and phonological templates. Functional linguistics cannot ignore the differential structural effects of these units – but neither can it ignore the functional principles that are at play. Thus, one must introduce an additional *structural* dimension (besides functional and social) responsible for the changes in replication. Structural stability is not directly dependent on prominence and entrenchment alone. Whether these properties cause normal or altered replication is determined by their structural environments. The structural environments, by determining whether linguemes in them undergo more likely normal or altered replication, can be understood as a *differential* phenomenon: morphs survive in prominent syllables and obligatory paradigms, but vanish in non-prominent syllables and non-obligatory paradigms. However, does this mean that we can interpret this as *differential replication*, i.e., *selection*?

Because of the structural nature of these factors, their accommodation into the Theory of Utterance Selection faces the challenge resulting from the premise that structural elements (linguemes) primarily function as replicators, not as interactors or environmental elements that would partake in the selection process. Thus, the TUS must either accommodate these structural effects within established concepts in the complex adaptive system of language or expand the theory. In Section 6.4, I will present arguments for how these effects can be interpreted under the TUS. However, before doing so, I will discuss in Section 6.3 how complexity of morph conditioning, which includes allomorphs, syncretic morphs and cumulative morphs can be analyzed along the lines of the TUS.

6.3 Complexity of morph conditioning from an evolutionary perspective

In the last section, I have argued that social factors do not fully account for the distribution and stability of certain ‘dysfunctional’ structures in language, since this ignores the association of complex inflection with prominent and frequent environments in utterances. To interpret this distribution from an evolutionary standpoint, it is necessary to first ascertain where morph types (allomorphy, syncretism, cumulation) and complexity of conditioning fit within the Theory of Utterance Selection.

To repeat Hull's (1988) and Mayr's (1978) statement above in the words of Blythe and Croft (2012: 271), evolutionary change is a two-step process: "the *generation of variation* in the replication process [i.e., altered replication], and the *selection* of some variants over others (also called differential replication)." Thus, morphs can be conceived as formal variants used to express a certain feature. In Chapter II, I have used the term 'morph variant' proposed by Haspelmath (2020) to refer to phonologically conditioned allomorphs. However, all types of conditioned morphs imply some sort of variation, a variation that is established across different conditioning contexts. While not a formal variant in the result, syncretic morphs can arise as variants of diachronic older, non-syncretic morphs, such as in the example of German articles, where the syncretic morphs *der* and *der* derive from the genitive feminine singular *dëra* and the dative feminine singular *dëru* respectively. Thus, there was a time in history where the last vowels [a] and [u] would drop in some environments and be retained in others, which would make syncretism a result of variation. What is propagated is however not the formal distinction (since *dëru* and *dëra* vanish, while *der* survives), but the semantic distinction between genitive and dative related to the form *der*. A similar persistence of semantic distinctions can be associated with cumulative morphs as well, which, once they arise as formal variants to non-fused morphs, retain their features; for example, German articles have retained the features for gender, number, case and definiteness in one morph, even if these features cannot be associated with certain segments. In English, only definiteness of the articles *the* and *a(n)* has survived.

One could therefore think that allomorphs, syncretic morphs and cumulative morphs are entities that have resulted from the process of *generation of variation* (altered replication), but that process has halted without undergoing further altered replication (phonological or semantic change).

Allomorphs, syncretic morphs and cumulative morphs are linguemes, because they are morphs, i.e., form-meaning pairings. The harder question to answer is whether *morphemes* (as opposed to *morphs*) are also linguemes. The Theory of Utterance Selection is ambiguous about this distinction but includes "anything from a phoneme to a morpheme to a word to a syntactic construction" into the definition of a replicator, i.e., lingueme. As argued in Chapter II Section 2.3, the dual structure of an underlying morpheme and derived allomorphs is not suitable for a usage-based approach to morphology and morphological complexity. The concept of a 'morph', according to Haspelmath (2020), circumvents the need to define the relationship between morphemes and allomorphs.

Because what is replicated is uttered language structure, the TUS would not need to consider the morpheme as a separate unit from the morph. In other words, what is replicated is units that are phonologically delimited, and this pertains to all morph types. The concept of a morpheme however implies a paradigmatic dimension to this structure: the replication of a morpheme would imply the replication of all morphs in complementary distribution with the same features. The TUS has not addressed whether paradigmatic relationships beyond the basic form-meaning mapping, i.e., variable elements in complementary distribution or in relationship to other paradigms, are replicated as well during the process of utterance selection. Thus, for a morpheme to be a lingueme separate from an allomorph there must be evidence that the paradigmatic dimension is replicated as a whole. Such paradigmatic replication may exist when altered replication affects the entire paradigm: for example, when all English past allomorphs (*-ed*, vowel alternation patterns) underwent a phonological reduction process or a semantic change at the same time. While there may be evidence for the latter, such as all morphs for past tense being recruited for protases of a hypothetical or counterfactual conditional such as in *If I **walked** on this bridge, ...; if you **drank** wine, ...; if he **lost** her wallet, ...*, there is no evidence that formal variants would undergo phonological alternation at the same time. Instead, allomorphy is the result of altered replication of one morph, and each new morph forms a lineage that does not influence the other lineages that have branched off. Furthermore, on closer examination, the interpretation for semantic change being evidence for *morpheme* replication instead of *morph* replication may be regarded as a change in the schematicity of a semantic category that is independent from the replication of the formal variant. In conclusion, it is more plausible to say that it is (allo)morphs that are replicators and not their 'underlying' morpheme.

Nevertheless, paradigmatic relationships between morphs need to be accounted for because conditioning contexts of inflection imply these relationships. In Chapter II Section 2.3 (Figure 2.5), a distinction has been made between conditioning and conditioned structure. The conditioning structure is not an abstract level of representation, but both structures co-occur in the same lingueme (the word). This means that when conditioned structure (morphs) is replicated, the conditioning structure (phonological environments, grammatical categories, lexemes) is replicated as well. For example, the English Plural allomorph *-ren* occurs only with *child*, and the English Plural allomorph of internal vowel change in *mice* occurs only in *mouse* and a couple other words such as *louse*. Since the morphs are not replicated independently from their context, one could argue that the conditioning and conditioned structure are part of the same lingueme. However, conditioning also implies a paradigmatic relationship between conditioning and conditioned structure. What distinguishes the instant replication of conditioned and conditioning structure in an utterance from the instant replication of two morphs that do not relate to one another in a word is that speakers establish a stronger connection between the first two structures. If we conceive of language from a connectionist or 'neural network' perspective, which is an idea that resonates with usage-based linguists (MacDonald & Christiansen 2002; McClelland & Bybee 2007; Traugott and Trousdale 2013; Joanisse & McClelland 2015), one can distinguish stronger and weaker connections between words and part of words. Most importantly, when two units that form a strong connection are replicated, the connection between these units is reactivated; one can say that this connection is replicated as well.

For example, the German circumfix to form past participles *ge- ... -en/-t* in ***ge**-schlag-**en*** 'beaten' and ***ge**-schnei-**t*** 'snowed' forms a strong neural connection between the prefix *ge-* and *-en/-t*, strong enough that *ge-* always needs to co-occur with one of those suffixes. In other words, *ge-* is *conditioned* by the occurrence of a past tense suffix. On the other hand, the connection between *in-* and *-ed* in English such as in ***in**-quir-**ed*** and ***in**-spect-**ed*** is not as strong. One could say that these two morphs are not 'relevant' to one another. The strength of these connections may be related to Croft's (2000) observation that "more inclusive linguemes as replicators often specify the structure of less inclusive linguemes that they contain" (p. 37). The type frequency of *in-* and *-ed* to occur in an English word is lower than the type frequency of *ge-* and *-en/-t* to occur in a German word; replication frequency of the more inclusive unit (word) determines the connection strength of the units included in it. On the other hand, the connection between *in-* on the one hand and *quire* and *spect* on the other hand may be stronger than between *in-* and *-ed* because the words *inquire* and *inspect* are both frequent replicators that strengthen the connection between the prefix and the stems. Following this connectionist perspective, one can say that morphological conditioning implies a strong neural connection between elements in words because they are part of more inclusive linguemes, and morphs that aren't conditioned by elements in words form weaker connections between them and other morphs because they form part of several, more independent linguemes. In sum, morphs that form a conditioning-conditioned relationship form strong connections by often co-occurring in linguemes.

Given this new contextualization of morphs and conditioning, we can interpret complexity of morphs (allomorphs and cumulative morphs except syncretic morphs) as the case where strong connections between morphs and word structure (phonological, morphological, semantic) come into being and are maintained despite these connections violating the one-meaning-one-form principle. While these connections might benefit processing of word structure by mutual hooks, it is difficult to learn (i.e., faithfully replicate) these connections as a first or second language learner. Failure to replicate these morphs may be due to the awareness that speakers must keep when replicating different linguemes of inclusivity; lexical conditioning requires the faithful replication of the morph *and* the more lingueme that includes it. The 'energy' invested in learning and faithfully replicating these linguemes requires the awareness of multiple connections that must be maintained, for the sake of conforming to social convention. Nevertheless, as stated before, social convention does not prevent change at every moment. It might explain why inflection is complex in utterances that are socially valued, but not the general distribution of types of conditioned inflectional morphs across a larger pool of utterances.

In fact, as this dissertation has shown, the distribution of phonologically, grammatically and lexically conditioned morphs differs in whether other principles facilitating communication are at play. As argued in Chapter II, phonologically conditioned allomorphy does not increase user complexity because the processes that lead to more morph variants might be beneficial (i.e., functional) for production or perception (such as assimilation or dissimilation). Grammatically conditioned morphs can be functional with regards to cumulation since it allows a speaker to include more information in one form – however, morphs becoming opaque and unpredictable in this process leads to deviation from many functional principles. Finally, lexically conditioned morphs do not seem to contribute any benefit for production or perception in contrast to other morph types. It is here were we expect that faithful replication of these morphs depends on factors unrelated to facilitating effects from properties of the conditioning context of morphs but depending on factors outside this conditioning-conditioned relationship, such as prominence and entrenchment.

In conclusion, the evolutionary perspective of conditioned morphs can explain why allomorphy and cumulation (and syncretism as an effect of it) come into being (as the result of generation of variation/altered replication without extinction of variants), but it cannot explain why some morphs are more stable than others besides conformity to convention. Within utterances, altered or normal replication alone cannot explain why morphs survive in certain structural environments; in other words: the tendency where morphs in prominent and frequent environments undergo normal replication and morphs outside those environments undergo altered replication constitute a contrast that can be described as 'differential'. But can this differential phenomenon be described as an instance of *differential replication* based on factors besides the interaction of speakers with their socio-communicative environment? In Section

6.4, the Theory of Utterance Selection is explored in relation to how it can accommodate the structural stabilizing effects of prominent and entrenched environments outlined in Chapter III.

6.4 Structural effects from an evolutionary perspective

In Chapter III, prominence and frequency have been identified as factors that facilitate the acquisition and processing of language structure, and in some cases also their stability. Facilitation and stability need to be kept apart; and I argue that only the latter is relevant for predicting the structural distribution of inflectional complexity. The reason is that structural facilitation does not always lead to structural stability. In evolutionary terms, whereas facilitation may or may not generate new variants, stability means that generation of novel variants is diminished. Facilitative traits of linguemes such as prominence and entrenchment might either lead to survival or extinction of variants which affects the lingueme lineages differently. Structures that are stable, on the other hand, stabilize the lingueme lineages. In Chapter III Section 3.4., the different effects of frequency on replication have been discussed. While both high token and type frequency increase the entrenchment of structures in the minds of speakers (and therefore facilitate processing), high token frequency of linguemes tends to erode phonological and morphological structure over time, whereas type frequency preserves it. This discrepancy can be also formulated with regards to entrenchment: deep entrenchment of morphological and phonological schemas stabilizes these schemas, whereas deep entrenchment of single morphemes and phonemes does not. Thus, it is the combination of a specific lingueme with a specific facilitative property that determines the structural stabilizing effect of that lingueme.

Likewise, it has been shown that prominence alone is not always a predictor of structural stability. It is stressed syllables that are stable and prevent extinction of included morphs. Morphs are also more faithfully replicated at the beginning or end of words, but the word boundaries are not as stabilizing as they are facilitative; as shown in Chapter III Section 3.3.1.3, word-initial onsets and word-final codas are more stable than word-internal intervocalic onsets, but the most stable consonants are 'word-internal onsets after obstruent codas' (cf. Ségéral and Scheer 2008). In sum, prominence and frequency alone do not predict lingueme stability, but lingueme stability is associated with prominent and frequent environments. This means that the specific type of lingueme must be accounted for when explaining the distribution of complex morphology across languages, pointing to the relevance of a structural connection between facilitating properties (here: prominence and frequency) and the structural phenomenon that is stabilized (here: complex morphology). The stabilizing effect is structural as it relates to specific linguemes (such as a syllable or paradigms) which mediate the facilitating properties. But where in the complex adaptive system of language can these structural effects be situated?

The first question is whether the normal replication of complex morphology in those environments is based on replication governed by functional principles. In this scenario, prominence and entrenchment as cognitive properties directly affect the replication of morphs; morphs that are prominent and highly entrenched are more likely to be faithfully replicated. However, the fact that the structural unit (syllable; morphological paradigm) mediates the effect between the facilitating properties and the normal replication of the morphs means that functional principles alone are not sufficient to explain the contrast between normal and altered replication of those morphs. Rather, it is the syllables and obligatory paradigms that are normally replicated, and the normal replication of morphs is the effect of it. This raises the question how to characterize this secondary process. As mentioned before, evolution is a two-step process: generation of variation and selection. So, under which process can the phenomenon be subsumed where replicated structure affects the change of overlapping replicated structure?

One interpretation is that the structural effects are effects only within the process of replication. Here, one replicator (e.g., stressed syllables) influences another replicator (lexically conditioned morphs). This calls for a more elaborated relationship between replicators/linguemes. In the following section, I will propose a framework in which linguemes exist at multiple levels, each with more and less inclusive units, leading to a complex overlap and interaction between them.

6.4.1 Linguemes at multiple levels

As Croft (2000) showed, linguemes can exist at different levels of inclusiveness:

> "[R]eplication of a syntactic construction requires replication of its component lexical items; replication of lexical items requires replication of their component morphemes; and replication of morphemes requires replication of their component phonemes. (Croft 2000: 11)"

This 'downward specification' is a core characteristic of emergentist behavior and shows the interlocking dynamics of inheritance. Based on the quote above, Croft regards morphology and syntax as a more inclusive lingueme than phonological linguemes. This is plausible when focusing on phonemes: phonemes are usually shorter than morphemes, and one could argue that the meaning components of phonemes are minor in comparison to a morpheme (as phonemes only express meaning in combination with other phonemes). But what about larger phonological structures, such as syllables, words and phonological phrases? Croft (2013: 43) regards the phoneme as ‘the functional lingueme [...] [in phonology] [...] [being] also a conceptual unit, though without semantic content.” whereas “the functional lingueme is characterized as a conceptual unit, usually associated with a component of meaning”. It is not clear whether greater strings of phonemes are considered linguemes. According to Croft, syllables and phonemes are emergent from larger gestural coordination patterns (Browman and Goldstein 1992) and schemas/templates (Bybee 2010; Vihman 2010 and Vihman and Croft 2007). However, I argue that because these templates and syllables are specifically structured and replicated across time and space, they can be also classified as linguemes. While there is a requirement that linguemes “are associated with a component of meaning”, one could argue that according to cognitive linguistics, larger phonological units also have meaning, even if a highly schematic one (Langacker 2008; Wilcox and Occhino 2016). However, one does not need to construct highly schematic meanings in order to account for concrete and prominent phonological patterns in the minds of speakers. Speakers are well aware of sound patterns, from as early as in the mother’s womb (Partanen et al. 2013; Lecanuet and Schaal 1996), and rhythm is embodied. There is no reason for these patterns to not be regarded as replicators, especially when comparing them to how melodies and rhythms are replicated in music. These patterns might be interpreted as having meaning, albeit a different meaning as the morphosyntactic schema with which they overlap. As Vihman and Croft (2007: 719) write: “phonology, like other aspects of language, must begin from the sound-meaning link that is central to the symbolic nature of language”. However, the overlapping phonological and morphosyntactic structure of words delimit the units included in them differently (e.g., syllables vs.

morphemes). And because both of these units are linguemes, one must posit two levels of linguemes (phonological and morphosyntactic) that can have multiple levels of inclusiveness.

One way to account for inclusivity within the phonological domain is to posit 'phonological templates' (Vihman and Croft 2007), a usage-based abstraction of the phonological structure of symbolic chunks such as words or phrases. Less inclusive phonological linguemes are syllables, which include phonemes. These linguemes are distinct from morphosyntactic linguemes (constructions, words, morphemes), but they co-influence one another as they are embodied in the utterance. Phonological structure can influence the change or stability of morphosyntactic structure, such as German umlaut, which is a phonological process that has led to variation and morphologization through reanalysis (Croft 2000: 128; Keller 1978). The opposite can also happen, where different affix types have led to the emergence of a predictable phonological template, like in Navajo: the erosion of distinctive phonological features in prefixes due to high token frequency of these affixes and the preservation of distinctive features in stems (which have low token frequency) has led to a bipartite phonological structure that contrasts a prominent (stem) and a less prominent phonological domain (prefixes) (McDonough 2003). This also has probably led to the emergence of so-called peg-elements, which are inserted for the faithful replication of the phonological template and are overgeneralized by children in first language acquisition (Chee 2017). So-called 'sesquisyllabic syllables' (words consisting of a light syllable preceding a heavy syllable) in several Southeast Asian languages are also an example of predictable phonological templates that evolved from prefixing structure (Butler 2014, Matisoff 1990, 2003, Sidwell 2000). The different lingueme levels and their inclusiveness can be illustrated in Figure 6.2.

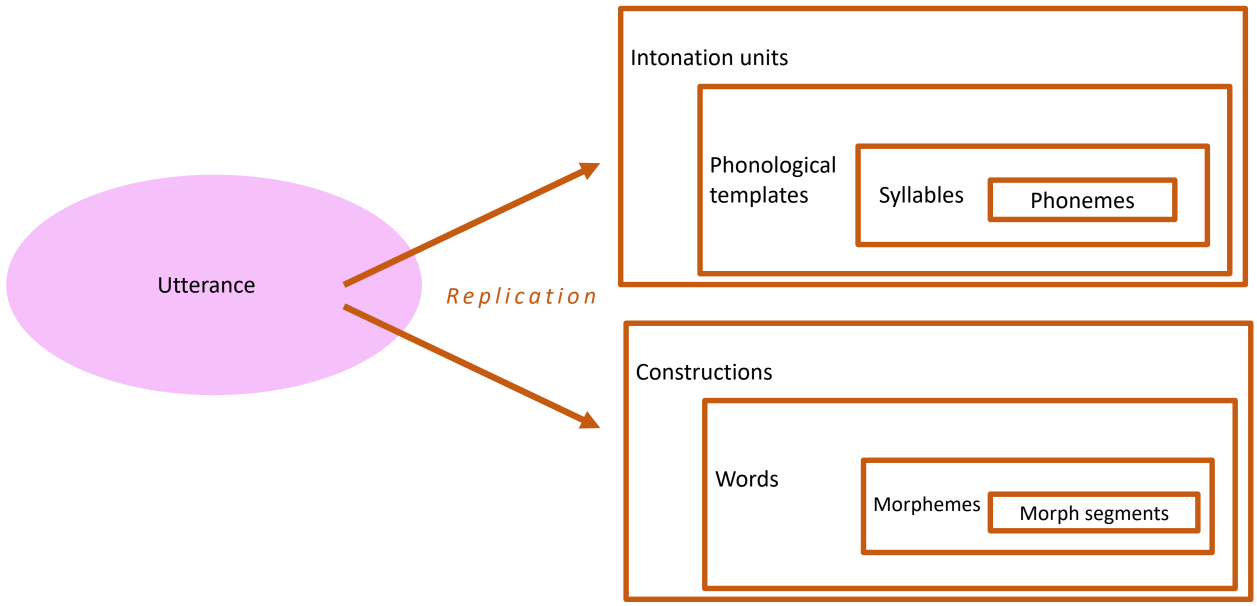


Figure 6.2: Inclusiveness of phonological and morphosyntactic linguemes. The parallel between intonation units and constructions on the one hand and words and morphological templates on the other hand is informed by Vihman and Croft (2007) but may be challenged. 'Morph segments' characterize sub-morphemic units such as alternating segments of stem morphs that exhibit conventionalized semantic features.

Figure 6.2 shows that utterance selection involves phonological and morphosyntactic replication. These lingueme levels exist parallel to one another as the utterance contains both phonological and morphosyntactic structure, and more inclusive linguemes specify the less inclusive linguemes. The distinction made between phonological and morpho-syntactic linguemes should not suggest that a phonological lingueme is replicated independently from a morpho-syntactic lingueme. Just as morpho-syntax is both dependent on phonology and semantic categories, so is phonology and semantics dependent on the morpho-syntax. But if these overlapping linguemes aren't replicated independently, why would there be a need to keep them apart?

The reason for doing so is that during replication, change of one lingueme does not always affect an overlapping lingueme in the same way. An example where change of one lingueme causes the change in an overlapping lingueme is given by Hare and Elman (1995): The dropping of English sounds in verbs led to the reconfiguration and merging of morphological classes expressing tense. An example where a phonological lingueme can change the morphology without causing change in the phonological lingueme was given in Chapter III Section 3.4.1.2. Blevins (2009) shows how Central Pacific Oceanic languages have a very stable CV pattern that causes loss of final consonants and thus allomorphy; therefore, an entrenched syllable pattern remains stable while causing changes in more included phonological and morphological linguemes. What this means from an evolutionary perspective is that the replication of a phonological template can influence the replication of morphology. Nevertheless, because the phonological template mediates this normal replication to the morphological template, one must account for this mediation either as an interaction between two replicators (in which case the structural effects are instances of multiple altered/normal replication processes) or as an interaction between speakers and the environment (in which case the structural effects are instances of *differential* replication). In any case, a more fine-grained analysis of replication is needed, one which can accommodate these biased effects.

6.4.2 Localization of structural effects in the selection process

As demonstrated in Chapter V, complex inflection is associated with prominent and deeply entrenched linguemes, indicating an interaction between these linguemes. The question is where this interaction can be situated in the process of selection, and how to interpret its effects.

One possibility is that this interaction represents a sub-process of replication. As argued in the last section, this interaction cannot be explained by altered/normal replication alone. The fact that functional/cognitive properties like entrenchment and prominence affect replication differently depending on the lingueme that is associated with these properties means that the normal replication of complex morphology is not merely the result of biased replication governed by these properties. Instead, overlapping linguemes mediate this bias for normal vs. altered replication.

One idea could be that differential replication of replicators caused by interaction of speakers with their environment causes further *differential* replication of included replicators. This is a plausible view when 'downward specification' is understood as a process. The stabilizing effects of normal replication occur differentially in this downward specification; i.e., certain linguemes are more likely to prevent change and extinction of linguemes included in them. According to this view, downward specification is not only a theoretical implication of replication but also of selection. From a complex-adaptive-system perspective, there is the possibility that smaller entities inherit the interactional dynamics of larger entities. However, from a usage-based perspective, it is questionable whether these processes exist separately from speakers interacting with their environment. Hull's (1988) General Analysis of Selection reinforces this view, according to which selection is an outcome of the interaction of interactors with the environment, selection being directed but without the aim to modify the system. The role of intention in selection is complicated; it has been argued in 6.2.1 that biased selection differs from biased replication in that the result of selection can be understood as fulfilling a preference between competing variants, whereas the result of replication can be directed but there is no preference based on competition. So, if the case is assumed where linguemes exhibiting facilitating properties are interpreted as interactors, causing a second process of differential replication, then one would have to argue that linguemes are the ones that ‘select’ between several competing overlapping linguemes.

According to the GAS, interaction is a process between the environment and the interactor, and selection a process between interactors and linguemes, causing differential replication. If structural effects are a type of interaction, this means that linguemes could either fulfill the role of an interactor or the environment. Let’s analyze the version where more inclusive linguemes are the interactor. This would imply that the environment is the speaker, and that the utterance or linguemes that are being selected by the speaker as a result of interaction with the environment causes a lower-order interaction between the utterance and speakers. This represents a recursive relationship of interaction between the environment, speaker, more inclusive linguemes and less inclusive linguemes. According to Hull (1988), multiple entities fulfill the criteria of interactors, at least in biological selection processes.

> “In order to function as an interactor, an entity must interact with its environment in such a way that some replication sequence or other is differential – the "relevant" replicators. Organisms are paradigm interactors. They are cohesive wholes, they interact with their environments as cohesive wholes, and the results of these interactions influence replication sequences in such a way that certain structures become more common; others rarer. However, many other entities also function as interactors – even genes. Genes have "phenotypes." DNA is a double helix that can unwind and replicate itself. In doing so it interacts with its cellular environment.” (Hull 1988: 409)

and furthermore:

> "Just as genes are not the only replicators, organisms are not the only interactors. Just as variable chunks of the genetic material function as replicators, entities at different levels of the organizational hierarchy can function as interactors. It does not take much to show that entities from genes to organisms can function as interactors. More inclusive entities are more problematic." (Hull 1988: 418)

In contrast, Hull is critical of whether there can be more candidates for replicators. Organisms have structure, but they do not pass this structure largely intact in successive replications. Rather, the genetic material in these organisms does. Even with asexual reproduction and a homogenous population, where "organisms can pass on their structure both directly and largely intact (...) the information that is transmitted is minimal when compared to the information passed on when the genetic material replicates." (p. 415).

Furthermore, the standard by which interactors cause replication raises doubts about the possibility of replicators also serving as interactors simultaneously. The rarity of entities that can fulfill the roles of both interactors and replicators could be a factor against considering linguemes to function in such a manner. Second, replicators aren't cohesive wholes, they are always part of utterances, and interconnected with other replicators (cf. Hull's 1988: 218 critique of the old conception that genes are "beads on a string" and not "hierarchically organized systems"). Utterances, on the other hand, can be regarded as cohesive wholes since they *contain replicators*. However, it is questionable whether utterances *cause* differential replication of replicators; for this to be possible, utterances would have to be to some degree external to linguemes, which is not the case (utterances are, among other things, structured sets of replicators, and as such the sum and combination of their linguemes). For example, although

Hull (1989) casts doubt on the idea, the cell in biology could be regarded as both an interactor (of differential replication of genes) but also a replicator (it passes its structure largely intact during replication), but genes are not the only content of cells. Utterances can function as complete replicators, in situations where the utterance itself is a distinct lingueme, as seen in fixed word phrases (such as the word 'goodbye' derived from the replication of the sentence 'God be with you'). If utterances were the only replicators, speakers would copy linguistic expression without creating novel combinations of linguemes. This is true for DNA strings: sometimes entire strings are copied, including every gene in it.

If we consider structural effects as a form of interaction, another scenario that arises is when utterances or linguemes are viewed as elements of the environment. Not in the sense that they are the environment of their included linguemes (this scenario would neglect the speaker completely), but part of the environment with which speakers interact. This interpretation has several benefits: it retains the idea where the speaker is the sole interactor and doesn't presuppose selection processes below the utterance. According to the General Analysis of Selection, there isn't any specification what functions as the environment, only that it can contain several elements or parts, and these are variable (Hull 1988: 466). In the Theory of Utterance Selection, Croft (2000: 38) specifies the environment as the social-communicative context. Nevertheless, infinite factors influencing interaction could contribute to the differential replication of linguemes, and the structural effects demonstrated in this dissertation are one among many. Regarding the perception of utterances and linguemes as environmental elements, it is important to note that this does not negate their function as replicators. The idea is that while speakers select utterances, they not only replicate linguemes, but speakers *interact with* these spatio-temporal units when they articulate and perceive them. Thus, speakers not only create the lingueme pool through selection and replication, but they find themselves inside it, surrounded by these linguemes, and these linguemes influence selection.

The sub-title of the dissertation contains the phrase 'prominent and frequent *environments*'. Prominent and frequent linguemes are not only parts of the utterance, but part of the environment that speakers interact with. While prominence and entrenchment influence the altered or normal replication of morphology (and create allomorphy), the structures that overlap with this morphology create an environment which can be described as causing differential replication: differential because these structural environments only influence the prevalence of normal replication of one variant vs. altered replication of other variants, they do not create new variants. Prominence may lead speakers to faithfully replicate a lingueme, but certain prominent linguemes determine whether this normal replication is lasting, leading to the differential distribution of complex inflection.

While stressed syllables and word boundaries are both prominent environments, stressed syllables influence normal replication of inflection longer than word boundaries. This has to do with the surrounding linguemes; word edges aren't replicators, but boundaries of replicators. If the preceding or following lingueme surrounding the boundary isn't faithfully replicated, this will affect the boundary. In contrast, if unstressed syllables surrounding a stressed syllable aren't faithfully replicated, this doesn't affect the stressed syllable in the same manner as word edges are affected by their syntagmatic environment. In any case, the effects between linguemes are part of the environment that *speakers* interact with, even if this interaction is the result of a former interaction that caused differential replication in the first place. Thus, the interaction of speakers with specific prominent and highly entrenched linguemes predicts which linguemes are stable (i.e., replicated faithfully) and which others aren't. This last scenario, whereby linguemes are not only replicators but part of the environment with which speakers interact, is visualized in Figure 6.3. The difference from Figure 6.1 is that the environment is expanded to include the utterance and the replicators, and that there is an additional interaction with speakers and this structure.

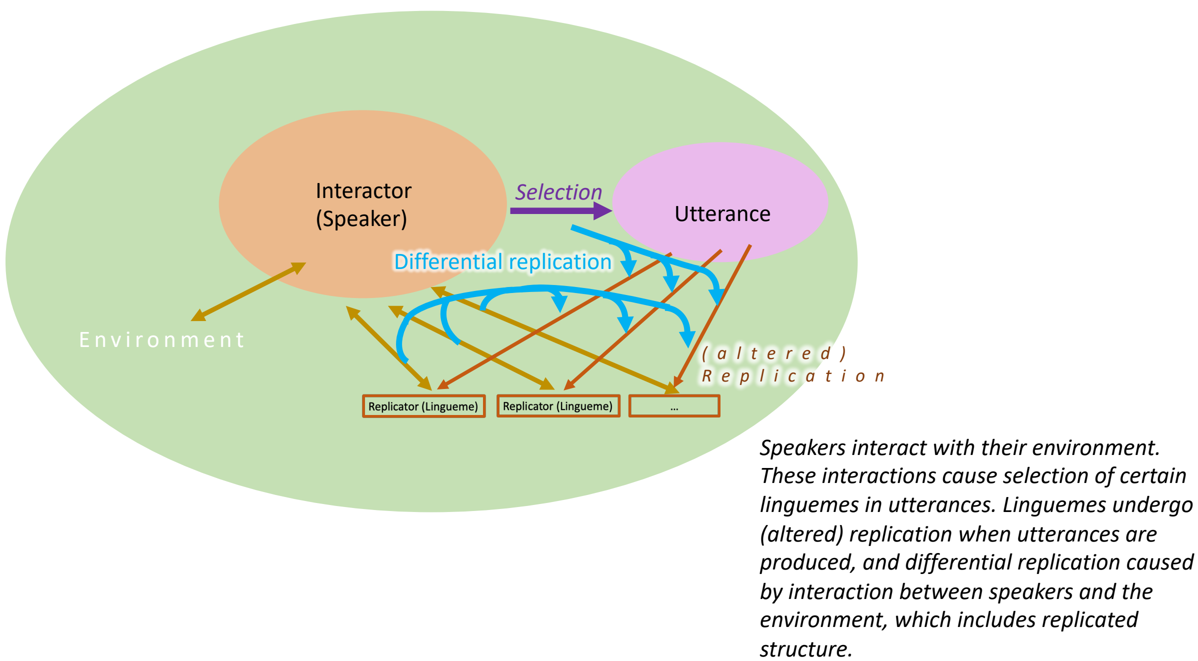


Figure 6.3: The Theory of Utterance Selection if the utterance and its replicators are considered part of the environment.

In conclusion, the reason why complex morphology is associated with prominent and frequent/entrenched positions is that linguemes are interconnected, and stable prominent and entrenched linguemes influence the stability of other linguemes. This stability does not directly originate from functional constraints determining normal vs. altered replication, but from the interaction of speakers with replicated structure, such as syllables and paradigms, which can be understood as an additional factor for differential replication.

6.5 Conclusion

This chapter provided a theoretical grounding to interpret the results obtained in Chapter V, and the premises outlined in Chapter II and III. It has been argued that an evolutionary perspective of complexity of conditioned morphs consists of the generation of variants that exhibit interconnected nodes in the mind of speakers; these variants are faithfully replicated despite violating the 'one form, one meaning' principle. The results in Chapter V show that structural effects that stabilize this complexity exist, and that these effects do not directly originate from functional constraints or social factors. In turn, the effects emerge as the product of replication of structure being aligned in certain parts of utterances and overlapping with other structural elements (linguemes). This structure-structure-interaction has been interpreted as a process derived from the effects that prominence and frequency has towards speakers and how prominence and frequency align with specific linguemes. Thus, functional-adaptive constraints still play a part in how linguemes undergo altered or normal replication, but specific linguemes (syllables, morphological paradigms) determine whether these effects are lasting, shaping the differential distribution of complexity in utterances. This differential phenomenon can be interpreted as differential replication caused by the interaction between speakers and replicated structure. This structure forms part of the speaker's environment and determines selection of variants as much as the functional or socio-communicative constraints do.

CHAPTER VII: CONCLUSION

7.1 Introduction

This chapter concludes the dissertation by providing a further discussion on the concepts and results presented in it. Chapter V showed that lexically conditioned inflection, and to a lesser extent grammatically conditioned inflection, is associated with prominent syllables and obligatory morpheme positions, which confirms the two main hypotheses according to which complex inflection is propagated in stable structural environments. In addition, Chapter VI provided an interpretation of this structural interplay from the Theory of Utterance Selection (Croft 2000), by concluding that the structural environment that contains complex inflection is part of the environment with which speakers interact, and this interaction can cause the survival of complex morphology. This chapter aims to give a more general discussion of the empirical, theoretical and practical consequences of this study. The conceptual frameworks established on the base of experimental and diachronic studies in Chapter II (Complexity of Conditioned Inflection) and Chapter III (Facilitative and Stable Structural Environments), as well as Chapter IV (Methodology) need to be reflected upon with regards to the results obtained by the typology. This chapter is outlined as follows: Section 7.2 discusses the theoretical and practical premises of Chapters II–IV in retrospective. Section 7.3 presents general implications for the concepts of morphological complexity and structural stability and their role in the language system. Section 7.4 reflects on the contributions of this thesis on the field of linguistics and suggests some options for further studies. Section 7.5 concludes this chapter and the dissertation.

7.2 Difficulty, facilitation and stability in structure

In Chapter I, the notion of language as a 'complex adaptive system' was evoked. This notion allows linguists to study the various aspects of language as connected and emergent entities. The premise of this study is that language is not only a complex adaptive system but incorporates some structures which are easier and other structures which are more difficult to master for speakers. Difficulty for speakers can have several causes, but one cause that associates specific structures with causing difficulty is structural complexity. Lexically conditioned inflection has been determined to be such a complex feature in language, that causes consistent effects of difficulties in speakers. On the other hand, language has *facilitative* structures. This dissertation has focused on both of these aspects of the language system and argues that structures associated with difficulty are embedded in structures that are facilitative, highlighting the interconnected aspect of linguistic units. Nevertheless, if facilitative structures were to always overlap with difficult structures, one could expect that the facilitative structures would win over time. The long stability of complex inflection in paradigms shows that this is not the case – facilitative structures do not reduce complexity, but rather perpetuate it. As such, a third factor, a factor which is purely structural, and sustains both difficult and facilitative properties of structure, must be posited.

The structural units that have been addressed in this thesis are syllables and morphological templates; and they influence the stability of inflectional complexity by being more prominent than its surrounding units (stressed vs. unstressed syllables; obligatory vs. non-obligatory paradigms). Stability of structures implies that some properties in these structures are facilitative, but facilitative properties do not imply that the structures associated with them will remain stable, and thus the structural aspect of facilitative linguistic units must be taken into consideration (besides the social constraint "conformity to convention") when explaining why some features are stable.

The complex lexical and grammatical conditioning of inflection does not need to be characterized as functional (there is little evidence to consider this type of morphology more effective than morphology that employs transparent form-meaning mappings), and as such its survival does not depend on benefits associated with these specific conditionings, but on its integration with stable structural environments which in turn might have functional-adaptive properties. Language-external forces such as stable discursive conditions as they exist in close-knit societies from residual geographic zones are not sufficient to account for why linguistic complexity grows and remains stable; the language-internal relationship to other structural entities is a smaller, but relevant factor as well. Illuminating the relationship between morphological complexity and stability entails bridging several gaps between linguistic levels and disciplines. This bridging involved a strategy to operationalize degrees of inflectional complexity through different conditioning contexts of morphs, and a strategy to operationalize the stabilizing potential of prominent and highly entrenched units. Finally, a strategy that allows the mapping of these two parameters was proposed, namely the construction of a spreadsheet with morphological positions and their potential for being prominent and obligatory.

This dissertation also attempted to bridge two linguistic fields that are not often connected: experimental research and typology. The typological variables in this dissertation result from generalizations of psycholinguistic and diachronic evidence regarding difficulty, facilitation and stability. This also applies to distinctions such as inflection vs. derivation and cumulation vs. agglutination; the cutting line was drawn based on semantic generality and decomposability, and psycholinguistic notions like declarative and procedural knowledge. While the results reveal the existence of a stable, ‘adapted’ complexity – i.e., complexity surviving in language because its structural ecological niches allow for it –, the multiple premises leading to the results in Chapter V and their evolutionary interpretation in Chapter VI need to be re-examined and discussed.

7.2.1 Conditioning of inflection as a variable for degrees of complexity

One of the concepts investigated in this dissertation was whether some morphological paradigms are more difficult than others. Lexical inflection served as a vantage point to establish a parameter of inflectional complexity that could reflect speaker difficulty through conditioning of inflectional morphs and paradigms. This solution involved generalizations of different approaches to complexity, running the risk of comparing incommensurable concepts. (Morphological) complexity remains a field with distinct measures that capture language- and user-based complexity. The idea of ‘added complex properties’ deriving from entropy, difficulty and description length (Table 2.12 in Chapter II) reflects this eclectic approach to determining levels of complexity. Nevertheless, difficulty for users has been the primary justification throughout this thesis.

Anderson’s (2015) parameter *Complexity of Allomorphy* proved useful for establishing morph types according to their conditioning, information that is usually found in grammars. In order to apply this parameter to all inflectional morphs, syncretic and cumulative morphs have also been included as instances of grammatically conditioned morphs. The clustering of PCI and UCI was also beneficial to increase the number of tokens that would constitute a category of lowest complexity of conditioning. Chapter IV and V have shown that there is also a distributional argument for the ranking of LCI, GCI and UCI. LCI occurs in languages that also have GCI or UCI[26], but UCI or GCI languages do not always have LCI. This distribution can be explained historically by acknowledging the more 'mature' status of LCI: LCI only evolves in languages that already have inflectional morphology; and if maturity is likened to higher complexity (Dahl 2004) this is an argument for LCI being more complex than GCI and UCI. However, the source of LCI need not be always GCI or UCI; LCI may also emerge as a pattern that is not over-generalized or is competing with another pattern across constructions. On the other hand, GCI is quite common across languages that have morphology (which are most of the languages in the world). It is not evident whether a language that has GCI also has UCI (this needs to be researched), but in many cases GCI derives from the fusion of two morphs that were formerly UCI (see Chapter V Section 5.3.3 for the distribution of fused vs. non-fused GCI morphs). In addition, morphs that are grammatically *and* lexically conditioned are rarer, and they tend to appear almost exclusively in prominent and obligatory environments. Thus, distributional properties of the sample align to some extent with the distinctions proposed in Chapter II. The distribution suggests that LCI is more restricted in its structural environment than other conditioned inflection types: LCI is mostly associated with prominence and obligatoriness, and UCI is least associated with these properties; GCI behaves like an intermediate category. The comparison with non-inflectional positions suggests that they are summarily less complex than LCI, somewhere between UCI and GCI. This might be due to clustering all non-inflectional positions as one, regardless of their function, which might neutralize the specific

[26] The Spreadsheet in the Appendix might suggest otherwise, such as Estonian or Ket. However, it must be reiterated that the designation of an LCI position is dependent on at least one morph in the position being lexically conditioned. In the case of Estonian, there are other morphs in the same position that are not lexically conditioned. The pattern in Ket is the actual exception and is not seen in other languages. However, Ket has quite productive nominal inflection that is not lexically conditioned, in which case the generalization made above must hold for all the inflectional domains.

complexity associated with certain morphs (e.g., stem morphs vs. valency vs. adverbial morphs). A parallel classification of lexically, grammatically and unconditioned *derivational* morphs would complement the study and give more insight into the distributive differences of inflectional and derivational complexity. The reason why this thesis has not undertaken a similar distinction among derivational categories is that these are less generalizable than inflectional categories. Also, they differ with respect to generality of their semantic features. Inflectional categories are all quite general in their semantics, and this property serves as a controlling variable to investigate the effect of different conditioning contexts of allomorphs, syncretic morphs and cumulative morphs. Furthermore, the difficulty that consists in faithfully replicating conditioned and conditioning properties in derivational morphology may be lower than in inflectional morphology, because the meaning of derivational morphs is more relevant to the meaning of the stem, and this relevance relationship might be facilitative in the memorization of words. For example, the stem and lexically conditioned derivation might be conceived as one entity, like in the word *per-ceive*. Instead, when speakers learn entire patterns for lexically conditioned inflection, they are less likely to conceptualize inflectional morphs and stems as one entity and must apply declarative and procedural knowledge to produce and retrieve the word form.

A relevant pattern that was detected as well is that the alignment of prominent syllables with complex morphs is more variable if the verb exhibits many positions with complex morphs. Languages that exhibit only one complex position usually have a fixed stress pattern which often falls on this position (as is the case with stressed stems in Dumi). Languages that have many LCI positions, such as Ket, show that every position can exhibit prosodic prominence, depending on the inflected word form. This association is independent from the pattern where LCI positions are closer to the stem than GCI or UCI positions. Thus, distributed complexity more likely entails distributed prominence.

The distribution of LCI in contrast to GCI and UCI across the languages and their alignment with prominence and obligatoriness suggests that the tripartite distinction formulated on the basis of Anderson's (2015) parameter *Complexity of Allomorphy* and modified to include only inflectional and all inflectional categories, is useful for a typological research. The implicational distributional relationship (LCI implies the existence of either GCI or UCI) reflects the psycholinguistic evidence and minimizes the risk of a circular argumentation according to which those morphs that are more likely expressed by stabilizing properties are more complex.

7.2.2 Prosodic prominence and morphological obligatoriness as stabilizing variables

The two independent variables that were contrasted to complexity of conditioned inflection are prosodic prominence and morphological obligatoriness. In Chapter III, several facilitating properties were reviewed with regards to their potential to be stable and stabilizing throughout time. One of the assumptions made in Chapter III is that stable structures must exhibit facilitative properties, but not every structure that has facilitative properties is stable. Another assumption is that stable structures can also stabilize other structures that they include. Thus, facilitative properties such as prominence and high frequency/entrenchment alone do not determine the survival and thus the distribution of difficult-to-learn morphology.

The relationship between stability and stabilizing potential is more intricate. Stressed syllables have been identified as being among the most stable phonological environments, but this doesn't mean that they never change. While stressed syllables will not lose their syllabicity (i.e., they do not lose their nucleus while being stressed), their replication might lead to change of other properties. The example in Chapter I (*vuestra merced* > *usted*) shows how univerbation leads to loss of stress in words. The first word, *vuestra* [ˈbwes.tra], loses its stress in the first syllable, and fuses with [merseð] to [us.ˈteð], which in turn inherits the stress from [mer.ˈseð]. Thus, stressed syllables do not necessarily retain the property [+stress] over time. Moreover, the change from [seð] to [eð] shows that stressed syllables do not stabilize their segments. However, they stabilize their nucleus, i.e., their syllabic integrity: it is highly unlikely that stressed syllables are dropped when they are stressed. Because of this, stressed syllables might stabilize the morphology that occupies this syllable. Thus, stability does not equal stabilizing potential. Rather, as argued in Chapter VI, the stabilizing potential is not inherent to structural units, but results from the *interaction* of speakers with their environment – the environment being here the replicated structure of a syllable. Certain units may be more stable than other units (i.e., stressed vs. unstressed syllables), but the stabilizing potential of stable units depends on (1) the interaction of the speaker with the unit by selecting it, (2) the overlap of this unit with other structural units and (3) the functional-adaptive ‘facilitative’ properties of the overlapping units. This complex speaker-structure and structure-structure interaction impacts the distribution of complex morphology across utterances. And these interactions can still be integrated in the framework of a complex adaptive system. This thesis does not argue that structure has more impact on language than speakers or usage do. The replication of more inclusive units affects the replication of more included units, but not vice versa. Thus, one limitation of choosing stressed syllables as stabilizing entities is that they cannot stabilize entire morphs that are larger than them. Furthermore, it was argued that further differential replication happens because speakers interact with the units they replicate, not because units interact with other units. As such, the causality of structural effects still aligns with the functional idea that embedded ecologies are affected by embedding ecologies, illustrating how geographic, social, functional, structural and – as this thesis has shown – inner-structural relationships form altogether a “cascade of ecological determinisms” (Mufwene 2014).

Shifting the discussion towards whether the distinctions of prominence and obligatoriness were effectively applied in this study, it needs to be noted that only a handful of stable units could comprehensively be mapped to the types of conditioned inflection. Issues arose for non-syllabic morphs, comprising either the onset or the coda of a syllable. Word boundaries were considered as an alternative and stricter criterion for prominence of consonantal morphs. While the effects of word boundaries on the distribution of types of inflectional morphs conformed to the expectations, the scarcity of purely consonantal positions (Chapter V, Figure 5.10) rendered the results non-significant. Because syllabic morphs were preferred when annotating the positions, this type of prominence (word edges) needs a separate investigation regardless of syllabicity.

Another issue that arose in this dissertation is the contrast in semantic prominence that occurs in verbs. Because inflectional categories semantically differ from lexical or derivational categories, it is difficult to account for semantic prominence as a variable. The most semantically prominent part of words are roots/stems since their meanings stand out as more concrete in contrast to the ones of affixes. One could interpret the distribution of inflectional complexity with regards to this type of prominence: LCI is the only type that is expressed in stems through stem alternation, and this conforms to the idea that more complex inflection is more likely retained in semantically prominent environments. However, this observation prevents a thorough examination due to the lack of non-lexically conditioned inflection in stems. It appears that stem alternation is almost always lexically conditioned. Stems alternate depending on their inflectional category and the phonological shape of the stem. However, one could imagine that inflectional stem alternation could be predictable, based on the phonological generalizations of stem shapes. The current sample did however not find any cases for this type of predictable stem alternation, although some argument could be made for Afroasiatic/Semitic languages where grammatical distinctions are expressed via predictable vowel sequences in stems. The only Afroasiatic language in this sample, Zuaran Berber, shows that the predictability is compromised as there are several lexically conditioned conjugation classes. Predictable inflectional stem alternation might not be very common and may not provide enough data to examine different conditioning types. Since inflectional stem alternation is always lexically conditioned in the sample, it has been decided to not research semantic prominence. When excluding stems from the investigation, the results still show that stressed syllables attract more complex inflection.

While obligatoriness has proven itself to be a significant factor in predicting the distribution of conditioned inflection types, it is a weaker factor compared to prominence. Obligatoriness as a category captures stronger entrenchment due to higher frequency of morphs but cannot provide exact counts. Obligatoriness and frequency might have a reciprocal relationship: obligatoriness might emerge because of high frequency, but it also can cause high frequency. Obligatoriness is one way to account for the stabilizing effect of high type frequency of morphs in obligatory positions. However, the structural effect is not due to high type frequency, but due to strong entrenchment of morphological schemas, as argued in Chapter III. It has been therefore decided to formulate obligatoriness as the potential for one position to be filled, strengthening the categorical representation of the morphological category/position.

This entrenchment of a structural position might prevent loss, as opposed to positions that do not exhibit an inflectional paradigm.

There is of course the need to find better strategies to capture this type of schematic entrenchment that leads to stability as opposed to attrition of structure. Phonotactic entrenchment also plays a role in retaining certain morphs, but the operationalization of strongly entrenched phonological schemas/templates needs to be improved. The distinction between obligatory positions that have zero morphs and positions that do not have zero morphs proved not significant as there are not many instances where a position is always phonologically filled in finite verb constructions. In general, there weren't as many obligatory positions in verbs as expected.

The scarcity of obligatory positions becomes clearer when excluding stems from the calculation. The issue with stems is that they imply obligatoriness, and as such it is hard to distinguish the effects of morphological obligatoriness from the effect of semantic prominence in predicting the distribution of complexity. Excluding stems renders the association between obligatoriness and complexity non-significant in the LCI sample (although significant in the total sample). Thus, stems contribute to the association between complexity and obligatoriness.

Finally, obligatoriness may develop from the fusion of distinct non-obligatory positions, resulting in GCI. This means that GCI might be rather the result of obligatoriness than obligatoriness being a factor contributing to the stability of GCI. The results show that cumulative morphs are associated with obligatoriness, indicating a historical fusion of frequent paradigms as the origin for the obligatoriness of the resulting paradigm. The relationship between obligatoriness and GCI is therefore not evident and can only be answered in a historical study that shows the chronological order of these developments.

As with the complexity variables, one can argue for the effect of additivity in facilitative and stabilizing structures. This study showed that the interaction of two or more stabilizing properties are a good predictor for the distribution of complexity. As expected, LCI is very often expressed by prominent, obligatory stem positions, comprising semantic and phonological prominence as well as high morphological frequency. These additive effects are also seen by the fact that prominence and obligatoriness more likely pattern with complex morphs in LCI languages than in GCI or UCI languages. The pattern suggests that the emergence and maintenance of highly complex inflection is more likely when there are more conditions that favor its replication. This means that the issues that arose while dissecting the different effects do not challenge the theory according to which more facilitative structures maintain less facilitative structures (complexity of inflection). In conclusion, this dissertation shows that the different functional-adaptive properties and structural units listed in Chapter III need to be studied separately (such as word boundaries or semantic prominence), but the convergence of these characteristics does not pose a problem for the general idea according to which a higher number of favorable conditions can maintain a higher number of complex properties.

7.2.3 Bridging operationalizations: methods revisited

The methodology reflects the challenges of bridging the different concepts in a practical manner. It presented the strategies used to obtain the results and intended to make this study replicable. In the following paragraphs, I discuss some issues that researchers need to be aware of when replicating the results or using these methods to study a similar research question.

Several preparations needed to be made to account for each variable and to achieve comparability across them. Most time has been spent on searching for languages with LCI, and to determine the status of the positions as LCI. In many cases, authors did not mention the restrictions for paradigms, and their conditioning had to be deduced from other parts of the grammars. On the other hand, searching for languages that have neither LCI nor GCI was a hard task as well. Most languages contain a certain degree of non-phonological allomorphy, because allomorphy itself is quite common. In addition, finding out whether morphs are lexically, grammatically or merely phonologically conditioned requires studying several sections in the grammars. The sampling of the languages involved accounting for several biases (see Chapter IV, Section 4.2). Furthermore, some linguistic geographic areas could not be included given the lack of complex inflection or inflection in general, such as in Southeast Asia.

Another challenging task was determining how to outline the morphological structure of verbs, and which morphs to include. The templates might not reflect the psycholinguistic reality of an utterance or a verb. They are merely a tool for mapping the complexity of morphs to their prominent and obligatory environments. The reason why positions and not morphs in isolation have been investigated is mostly practical: it facilitates the designation of obligatoriness. Obligatoriness does not pertain to specific morphs, but to categories that are always expressed by morphs in a specific position/paradigm.

However, capturing the effect that specific morphs and paradigms have on language speakers was not neglected. Like the classification of complexity and facilitative properties, the distinction between inflectional and derivational morphs follows a functional, not formal approach. Relevant is not which grammatical category they belong but whether they are semantically general. Therefore, the search for inflectional elements was not restricted to affixes but it incorporated frequent words that have general semantics. These include auxiliaries, which were encoded as either derivational or inflectional, based on the meaning of their stem.

On the other hand, the non-inflectional positions in the templates include unbounded adverbials that express derivation and aktionsart. Section 4.4.2 in Chapter IV discusses how the morphs were selected and their positions determined. The specific criteria to determine the templates resulted in some analytic languages looking more (poly)synthetic, as is the case with the template of Standard French, which has 15 positions. The liberal solution to include unbounded elements brought the benefit of increasing the number of datapoints for the different types of inflection, amounting to 327 positions in the LCI sample alone.

Flowcharts were a tool to determine various choices of annotation; they are based on functional premises, but these premises were developed in the process of data gathering, morph classification and definition of a position. The determination of morph types, prominence and obligatoriness also evolved throughout data gathering. This might have resulted in classifications specific to this study, which can be changed or made more precise in further studies. The flowchart in Figure 4.4 (Chapter IV) is a solution specifically designed to account for the distribution of prominence categories found in this study. It does not represent a solution to classify prosodic prominence in general. For example, it might be appropriate to formulate separate categories for stress, tone and weight. Likewise, obligatoriness in Figure 4.6 was assessed by accounting for both phonological (overt) and semantic obligatoriness (non-overt), where a better solution would be keeping these as separate variables.

The annotation sheet in the Appendix reflects the empirical core of this dissertation. As mentioned above, most time in this study was spent on gathering data, and gradually building the sheet. During this process, some categories needed to be redefined, and positions merged or split. The annotation sheet is a useful source to obtain replicable results, but it also doesn't tell the entire story. While there are explanations for why inflectional positions have been classified as LCI, GCI or UCI, little information is provided on the non-inflectional positions and how these were determined as such (besides the flowchart in Figure 4.2). Furthermore, sometimes the assignment of prominence has been deduced from rules mentioned in grammars, which means that examples that demonstrate stress placement are only occasionally provided. Nevertheless, the annotation sheet comprises sufficient information for readers to trace back the annotation choices to specific pages in the grammars. In conclusion, this thesis enables replicability of the results through the spreadsheet and flowcharts presented in Chapter IV. These tools can be adopted and refined to research further questions on inflection, prominence and obligatoriness.

7.3 Implications

After discussing the categories established in Chapters II, III and IV based on the results obtained in Chapter V, this section proposes more general implications of this study. The results are relevant for the theory of complexity and structural facilitation and stability. Thus, this section comments on the ideas introduced in Chapter I, with regards to the results obtained in Chapter V.

7.3.1 Complexity

Although the research in the last decades has aimed for more objective, language-based measurements of complexity, this dissertation shows that it is possible to generalize and operationalize user-based complexity, complexity that reflects speaker difficulty. Nevertheless, Chapter II also argued that language-based measurements can be related to user-based measurements. Since this dissertation investigates reasons for the stability and distribution of a linguistic phenomenon, the speaker-based approach was prioritized; the reason is that only complexity that has effects on speakers will also interact with other effects – the facilitative effects discussed in Chapter III. While only morphological complexity has been targeted, it is possible to investigate the notion of complexity in other levels of language, such as phonological, syntactic or semantic complexity. Not everything in language is directly functionally motivated, so the task for functional linguists is to uncover the elements that mediate the motivations among users through structure and time. Source-oriented approaches trace the motivation to a specific construction in history where the motivation is transparent, whereas result-oriented explanations identify the different functional-adaptive constraints that are still at play in the phenomenon under study, even if the transparent motivation has been lost. The study presented in this dissertation argues that hard-to-learn phenomena in language cannot be cancelled out when finding these constraints or motivations, but that there are structural interactions for why some parts in utterances are more challenging for speakers than others. Speaker-based complexity is not marginal but can be at the center of the language system.

7.3.2 Structural facilitation and stability

The results can be accommodated with the ideas elaborated in Chapter III. One claim is that the stabilizing effects of smaller structural entities can be studied. Speakers select utterances and replicate structure in those; however, they have less control on how the properties of these structures interact with other structural properties. This structural interaction must be accounted for when asking for the distribution of smaller complex units in languages. The question is whether this interaction has a significant impact on the language system. A relevant implication from the results is that the mapping of several stabilizing properties increases the effect of these structures on others, which motivated the theory presented in Chapter VI. The more stabilizing properties overlap, the more likely properties that are less facilitative (such as lexically conditioned inflection) are retained. While these additive effects have not been studied thoroughly, they point to the existence of stabilizing properties on different structural levels, as proposed in Chapter VI, Section 6.4.1('linguemes at multiple levels'). This confirms the conception of an utterance as a bundled entity where more and less facilitative structures diverge and overlap.

This dissertation answered that the impact of structural effects exists, but it failed to research how strong each of these effects are. While the results show that stress is a better predictor than obligatoriness for the distribution of complex inflection, it is questionable whether these two variables can be compared, given that they operate on different levels with a different dependency on other properties of morphs (obligatoriness derives from stems and fused morphs; stress does not). This points to the general question whether there is commensurability between stabilizing structures across levels. The general principle is that larger units can impact the behavior of smaller units, not vice-versa – a non-reductionist principle that is compatible with functional and evolutionary approaches to language.

While the effect of stabilizing structures has been shown in the distributional properties of conditioned inflection types, this dissertation also cannot fully address the distinction between facilitation and stability. Chapter III suggested that prominent syllables and entrenched morphological schemas are stable, but in order to show that prominence and obligatoriness are better predictors for the distribution of complexity in words they must be compared to other prominent and frequent structures, which need individual operationalizations. This dissertation used this distinction to select promising variables to predict this distribution, and this proved useful, but it was not shown whether other structures that are facilitative for acquisition are also stable across replication events. Again, the general principle is that facilitation is the precondition for stability, not vice-versa. The effects of facilitation and stability, and whether structure itself (syllables, segments, morphemes) is a factor that determines the latter, is something that will have to be addressed in further studies. A diachronic study on the persistence and change of different stable structures can be conducted to shed light on the different effects.

7.3.3 Relationship between complexity and functional-adaptive constraints

This dissertation has suggested and shown that the relationship between functional-adaptive properties (like prominence and entrenchment) and complex structures is a passive, non-compensatory one. This can be explained as follows: some properties in structure facilitate communication; as a result, speakers are more likely to produce and perceive these structures in utterances. Speakers do not care about the structural interaction between these structures; they merely produce utterances based on already replicated structure. One can say the emergence of complexity may in fact just derive from *altered* replication processes, whereas the stability is due to *normal* replication that is favored because speakers interact with replicated structure. Linguists who attribute some function to (morphological) complexity overlook the fact that any structure in language contains more and less facilitative properties. Less facilitative properties like lexical conditioning do not have a hidden functionality that needs to be uncovered, instead, one must consider the structural environment in which they are embedded. Thus, the relationship between facilitative properties and non-facilitative properties such as inflectional complexity follows the same principle as the one between facilitation and stability: facilitative properties support non-facilitative properties, not vice versa. Furthermore, this support is not directed, but emergent: facilitative properties do not accommodate themselves in verbal utterances so that speakers can process difficult features more easily. Instead, prominent and frequent environments are contexts where complexity survives, and if complex features are not supported by these contexts, they would likely vanish over time. This loss would not impact the functional-adaptive constraints that are always present.

7.4 Contributions to the Field of Linguistics

This portion examines the connection between this study and the field of Linguistics, emphasizing its contributions and suggesting potential avenues for future research. Of course, it is hard to picture which aspects of this dissertation will be useful for the future, and how it could impact linguistic theory. The first step is to locate the novel ideas and explain how they elaborate on, or challenge established frameworks. The second step is to propose possibilities to research these new avenues.

One contribution of this dissertation is the formulation of typological variables based on experimental and historical evidence. The novelty in this approach consists of the deductive process by which experimental evidence is generalized to classify morphs, prominence and obligatoriness, and enable comparison. Typology usually works with comparative concepts to explain structural distributions; yet these comparative concepts include categories that might not be processed similarly. The method applied here, if propagated, could foster a closer connection between experimental research and comparative linguistics. As Norcliffe, Harris & Jaeger (2015) observe, the field between psycholinguistics and typology is growing.

A more specific typological contribution of this dissertation is the idea that classifying inflection according to their conditioning is relevant for researching the effects of difficulty on speakers. This not only includes allomorphy, where the concept of complexity/difficulty being dependent on conditioning was defined by Anderson (2015), but also other inflectional morphs where the form-meaning matching is not symmetrical, such as syncretism and cumulation. This dissertation also has propagated the term 'morph' (as defined by Haspelmath 2020) in order to depict a form-meaning entity that has not yet been specified in relationship with other form-meaning entities (other allomorphs and paradigms) and conditioning environments. As such, the concept of morph positions is proposed instead of morpheme positions. Morphs can be conditioned or unconditioned, and their notion can also account for sub-morphemic segments of stems that were relevant for this study (stem alternation). Thus, the discussion of morphological complexity requires a level of analysis smaller than the morpheme, and this thesis intends to propagate the concept of 'morph' to work with. Furthermore, this dissertation propagated Bybee's (1985) way to distinguish between inflectional, derivational and lexical morphs, by using the concept of semantic generality and paradigmatic distinctions. While the continuum between these notions is acknowledged, typologists must draw practical lines between them to apply them for language-comparison.

In addition, this dissertation has provided an evolutionary framework to compare different structural levels of words, such as between morphology and phonology. This opens the possibility for further studies on the relationship between semantic, morphological and phonological co-occurrences across languages, fostering the connections between linguists from different disciplines. Typological studies are usually restricted to one linguistic level due to the specialization of the linguists, but language as a complex system shows connections among them that need to be discovered. The framework outlined in Chapter VI (linguemes affecting other linguemes) can be transferred to other studies. One option is to investigate whether facilitating structures stabilize phonological complexity as well. Easterday (2017: 453ff.) noticed a positive correlation between syllable complexity and degree of synthesis (ratio of morphemes per word). The fact that these two complex properties overlap might speak against the theory proposed here; however, it must also be studied whether complex syllables exhibit facilitative properties themselves. Complex consonant clusters or rare phonemes may contribute to prominence/salience, which in turn is a facilitative property. Following this field of study, it would of course be worth investigating how complexity of inflection or morphology in general aligns with complexity of syllables.

The interpretations of the results in this dissertation would not be possible without a theory of language change. The evolutionary framework, more specifically the Theory of Utterance Selection (Croft 2000) was chosen to interpret the stability of complexity. The reader might have missed an additional diachronic investigation of the research question, and it could have been added to the synchronic distributional analyses. Unfortunately, time constraints prevented such a study. The connection between synchrony and diachrony could be established by assuming that *differential distribution* (complex inflection is mostly found inside prominent and frequent environments of utterances that outside of those) is the result of *differential replication* (speakers interact with prominent and frequent environments which cause more stable morphology in those environments than outside those environments). A future diachronic approach to the research question would involve identifying the chronological relationship between the emergence of stabilizing structures and the life cycle of morphological complexity. This study has suggested that lexical conditioning is a mature phenomenon that doesn't have a functionality of its own in contrast to non-lexically conditioned inflection, so it is expected that complex inflection emerges before the prominent and frequent environments that support it cease to exist. This research may not require a comparative approach and would also work within a specific language family.

Arguably, this dissertation expands on the notion of structural interaction from an emergentist perspective. Emergence in language is often invoked when explaining the development of categories and structural elements. Language as a complex adaptive system is conceived as emergent. The notion is often used to characterize the coming to being of phenomena; the development of systems that are self-regulatory (such as ‘stigmergy’; see Heylighen 2016) is often missed. This dissertation researches the possibility of these systems by pointing towards possible feedback-cycles between speakers and structure. Thus, it might be the case that the language system may develop further cycles below the interaction between speakers and utterances. However, as argued in Chapter VI, language structure is not self-regulatory, but always needs the speaker (an interactor) to cause its differential replication. The fact that linguemes influence other linguemes is not seen as a life cycle of its own, but this influence is always tied to the ecology in which speakers interact with their environment. Nevertheless, there is still work needed on the relevance of structural effects within a functionalist framework. This thesis has provided a small step towards the substitution of formal approaches on structural interaction by an integrated complex-systems approach.

This dissertation has also addressed the question of functionality in language. By researching complexity as user difficulty, the relevance of phenomena that evade functionality was brought into focus, as co-equal components of complex adaptive systems. Thus, entropic ‘noise’ – structure that gets left behind in optimization processes – is as well part of the complex adaptive system of language. While the explanatory power of trajectories in change lies in functional-adaptive explanations, the pace at which change occurs is also dependent on properties that are not functional-adaptive. This means that for example, the abolition of entropic properties like allomorphy, syncretism and cumulation in language might be stalled by environments that are otherwise functional-adaptive. As much as linguistic phenomena can be explained by functional constraints, one could always find arguments for why a specific construction is not optimal. Acknowledging the two parts of language might incentivize research on the 'dark matter' in language, the irreducible parts which have been either ignored or have been unnecessarily inflated with functionality. The solution is however not to be satisfied with purely source-oriented explanations that explain the development of structures by language-specific processes. What is needed is capturing the nature of this entropic field of language and identifying common factors for how it emerges, changes and is stabilized. Thus, new studies could investigate how functional-adaptive constraints are themselves constrained by the limits set through the structural or entropic environment that speakers interact with.

A more practical contribution of this dissertation is the data displayed in the spreadsheet. Researchers might be able to use this sheet to study morphologically complex languages. This information is a valuable addition to larger databases such GRAMCATS (Bybee, Perkins and Pagliuca 1994) or as the recently published GRAMBANK (Skirgård et al. 2023), with inflectional morphs and conditioning further specified, and subsumption into positions. The data can be used to investigate further interactions of structural elements. For example, one could research the distribution of conditioned inflection in languages in relation to the most commonly used constituent order in these languages. The distribution of conditioned inflection types could also be further investigated with regards to prefixing vs. suffixing morphology, or to number of morphological positions. The implicational relationship of the distribution of LCI, GCI and UCI (or phonologically conditioned inflection) is still in need of investigation.

There is also the need to annotate other types of prominence, such as word boundaries or syllable weight. Weight was used as a sidestep strategy when none of the morphs in a position were syllabic or when the languages did not have stress or tones. Keeping these types of prominence apart (also with regards to stress in tone languages) and analyzing them individually might uncover the impact of each of them on other structural levels.

In sum, this study enhances the understanding of operationalization of psycholinguistic variables in typology, the concept of conditioned inflection, morphological complexity and the morphology-phonology interface; furthermore, a diachronic, evolutionary view reinterprets and expands on the ideas of structural interaction and functionality. Finally, this dissertation provides enough data in the spreadsheet to conduct further studies with similar research questions. It was attempted to provide examples for future studies in each of these linguistic realms, but the recommendations are of course not exhaustive. One could imagine that this type of research inspires linguists to think generally about where structural interactions are located in the complex system of language and how to integrate different methods to uncover these interactions.

7.5 Conclusion

This chapter has provided a general discussion and interpretation of the results. The results show that complex morphology such as lexically conditioned inflection is associated with prominent structure such as syllables bearing stress and prominent tonal contrasts, and highly entrenched environments such as obligatory morpheme positions. These results have been obtained by bridging theoretical frameworks to operationalize distinct categories used to reveal the relationship of complexity of inflection, prosodic prominence and morphological entrenchment. A framework was provided that shows how the stability of complexity results from the stability of structures interacting with their structural environment as well as with speakers. The intertwined relationship of stabilizing structures can be used to explain the distributional relationship among structural elements. This sheds light on the implications for functionalist theory: not every linguistic phenomenon has a hidden function, and neither are there phenomena that are fully accidental/arbitrary; instead, more and less facilitative properties reside within every element of language; sometimes several properties overlap and cause features that are more challenging for speakers to persist. Excavating both functional and non-functional aspects of language should be studied together, as exemplified in this dissertation. Further studies will show whether the generalizations are also reflected in diachrony, i.e., whether the change in facilitating and stabilizing structures leads to a change in structural complexity.

Appendix

The following table describes the categories of the columns and the abbreviations used in these.

Column	Colum Group	Category	Abbreviation explanation
A	Morphological position, categories, morphs and features/values	Position Number (Pos.)	Suffixes: positive numbers Prefixes: negative numbers 0, 0.1, 0.2, 0.3 etc.: stem, stem segments and infixes
B		Functions	General functions: LEX: lexical morphs (main and compound stem) V: Valency T: Tense M: Mood A: Aspect P: Person N: Number G: Gender C: Clusivity CL: Classifier PO: polarity E: Evidentiality HO: Honorific NP: Noun phrase ADV: adverbial affixes (a sundry category where aktionsart, pluractionality, discourse, manner, and derivational morphs except valency are placed)
C		Categories	The language-specific categories (such as iterative, thematic, reflexive). For indexation, the grammatical role is mentioned (subject, object etc.)
D		Morphs and features/values	The (allo-)morphs and their features are listed for the given category.

E	Conditioning (Cond.)	Values	L = Lexical conditioning of some morphs (can be additionally grammatically and phonologically conditioned) G = Grammatical conditioning of some morphs (never lexically conditioned, can be phonologically conditioned) U = Unconditioned. Neither lexically nor grammatically conditioned. Can be phonologically conditioned. X = Non-inflectional position; conditioning not assessed
F		Explanation	Explanation and reference for the choice of annotating a specific conditioning type for the position.
G	Prominence (Prom.)	Values	**yes**: position is always prominent (syllabic morphs are always stressed or have highest variety of tones; consonantal morphs are always part of syllables that are stressed and may occur at the beginning or the end of words) **possible**: morphs in the position are prominent in some constructions **no**: morphs in the position are never prominent Values in parentheses represent monoconsonantal positions.
H		Explanation	Explanation and reference for the choice of annotating a specific prominence value for the position
I	Obligatoriness (Obl.)	Values	**yes:** position is always filled **(yes):** paradigm is obligatory and occurs only in this position but exhibits zero morphs **no:** position is not obligatory; zero morphs appear in most of the forms
J		Explanation	Explanation and reference for the choice of annotating a specific obligatoriness value for the position

AGUARUNA (LCI sample); Overall (2017)

Prominence: stress/pitch.
"The default root accent surfaces on the second vowel of the underlyingly disyllabic roots. If a verbal word contains no lexically-specified accent, whether in the root or a suffix, then accent falls on the root by default. (...) When an accentuated vowel is elided, accent shifts to the next accent-bearing suffix, or again defaults to the root if there is no such suffix." (p. 99) "A few suffixes and morphological combinations can trigger accent effects without actually taking the accent themselves." (p. 100)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-1	V	Causative	i/a/ɨ (causative). Prefixed vowel, dependent on verb (not productive) (302)	X		no		no	
0	LEX/A	Root, perfective, imperfective, potential, durative	Final vowel alternates according to perfective, imperfective. (273-278)	L	"A number of morphological changes and phonological rules interact to define a total of five conjugations. The conjugations are based on variation in the root when forming the unmarked root form. The morphological phenomena that define the conjugations cannot be explained in purely phonological terms." (272)	pos.	"The default root accent surfaces on the second vowel of the underlyingly disyllabic roots" (99) which is the final vowel in many cases.	yes	
1	V	Causative	miti(ka) (causative) (303)	X		pos.	No examples found where this suffix are stressed.	no	
2	V	Applicative, detransitivizer	tu/hu (applicative) (306) na (detransitivizer/non-valency changing) (307, 307) ki (S>A transitivizer) ka (S>O causativizer) pa (A>S detransitivizer) (311)	X		no	No examples found where these affixes are accented.	no	
3	V	Reflexive/ reciprocal	ma/mama (reflexive) (311) nai (314)	X		pos.	See 313 for reflexive stressed: tumámiuwai tu-mama-i-u-ai say-REFL-LOAF-REL-COP:3DECL 'he said that he (hadn't eaten beans)'. See 314 for reciprocal stressed: timaʃnáyainawai timaʃi-nai-ina-wa-i comb-RECIP-PL:IMPFV-3-DECL 'they are combing each other's hair'	no	
4	P/N	Object	hu/tu (1sg) hama/tama/pa (2nd) hama/pa/hatu/tama/tVpa (1pl) Ø (3rd person) (315)	L	"[V}ariants are (…) lexically conditioned" and depend on verb class (316).	pos.	Can be stressed. See p. 315 for 1sg obj dakumhúkta dakuma-hu-ka-ta copy-1sg.OBJ-INTS-IMP 'Take my photo'.	no	Third person object marked zero and these suffixes appear only in transitive clauses.

AGUARUNA (cont.)

Pos.	Catego-ries	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
5.1	A	Perfective	Ø/s/ha/há/kí/ká/í/a/sá/sa/á (Perfective) (274-278)	L	Lexically conditioned only in perfective (since expressed through different aktionsart allomorphs). "The Aktionsart suffixes are semantically vague, and it is very difficult to pin down a meaning for each one." (292)	pos.	All of the Aktionsart suffixes except i(ni) 'low affectedness' can take accent. (292)	no	Usually present, but there are zero forms when there are tense suffixes or nominalization suffixes.
5.2	A/M/N	Imperfective, imperfective plural, potential, durative	Ø/a/á (Imperfective) (274-278) (i)na/ína (Imperfective plural) (82) mai (Potential) (299) ma (Durative) (301)	U	No conditioning found except phonological (278)	pos.	All the imperfective, potential and durative markers can take a high tone. (see paradigms and examples (82; 274-278; 299-300)	no	Usually present, but there are zero forms when there are tense suffixes or nominalization suffixes.
6	PO	Negation	tsu (Negative present and future tenses) tʃa (negative other) (324)	G	Allomorphy conditioned by tense (481)	no	Negative is never stress, instead, it "causes accent to fall on the root in all forms" (100)	no	
7	A/N	Perfective plural	aha (perfective plural) (289)	U	No conditioning found. Can only appear in perfective roots (289).	pos.	Can bear stress, see example 140 tsupi-ká-ŋ-tat-u-i cut-INTS-PL-FUT-3-DECL ‘they will cut’ (83)	no	

AGUARUNA (cont.)

Pos.	Catego-ries	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
8	T/M	Narrative past (nominalizer), (indefinite) future subject (nominaliser), desiderative, imperative, jussive, hortative	haku (narrative past) (334) tinu (findefinite future subject) (356) tata (desiderative) tatuina (desiderative plural) (354) ta (Imperative) (350) kia (familiar imperative (351) ti (Jussive) (353) mi (Hortative) (360)	U	No conditioning found in the respective descriptions.	pos.	Can bear stress, see example 140 tsupi-ká-ŋ-tat-u-i cut-INTS-PL-FUT-3-DECL ‘they will cut’ (83)	no	
9	T/M/P/N	Subject, object, present/definite future, immediate future, immediate past, recent past, intermediate/distant past, remote past, apprehensive, prohibitive, subordinate verb (365)	Different TMA-Person-Number allomorphs (365), e.g.: Non-past declarative: ha (1sg) hi (1nonsg) mɨ (2sg) humɨ (2nonsg) (361); Past declarative: ha (1sg) hi (1nonsg) umɨ (2sg) uhumɨ (2nonsg) (362) wa/u (3sg non-past declarative/polar interrogative/exclamatory) Ø (3sg person other) mɨ̃ (3sg declarative recent past) ma (3sg recent past other) yi (3sg declarative remote past) ia (3sg remote past other) ti (3sg immediate future) (364)	G	Allomorphs for person are conditioned by several TMA categories (365).	pos.	Can bear stress, see example 201 uti-t-há-i 'fetch-IFUT-1sg-DECL' 'I will fetch (it)'. (101)	(yes)	"The major suffix groups at level II, namely tense, subject and mood, are obligatory" (270). Zero marking in only one category (non-past third person 'other') (364), zero-syllabic.
10	M	Indicative (declarative, counter-expectation, narrative, speculative), interrogative (polar interrogative, content interrogative) (suppression of apocope, clause contains interrogative word), tag question, exclamative	i (Declarative) hama (Counter expectation) tuwuhamɨ̃ (Narrative) tai (Speculative) ka/Ø(if marked elsewhere in the clause) (Polar interrogative) suppression of apocope (Content interrogative) api (Tag question) Ø (Exclamative) (366)	U	No conditioning found (366).	pos.	Can bear stress, see p. 101 for example 200 where declarative is marked with an accent due to elision of preceding vowels: [tipisuí] tipi-sá-u-i lie.down-ATT-REL-COP:3:DECL ‘he lay down’ (101)	(yes)	"Mood is obligatorily marked in finite clauses, and all clause types are marked with suffixes in [this] slot [...] except imperative, marked in [position 8], and some interrogative types which are marked on constituents of the clause." (330) However, some forms are zero, like exclamative and polar interrogative.

AYUTLA MIXE (LCI sample); Robero-Mendez (2009)

Prominence established on stress/pitch. "In verbs (...) the last syllable of the verb stem is stressed. The verb stem is defined as what is left after the inflectional morphology is removed" Footnote: "The leftmost limit of the verb stem is defined by the last verb root in it." (80)

Pos.	Catego-ries	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-11	P	Subject/Object person markers	Independent: Ø (1,3, 3'>3) m (2, 2>3, 3>2) n (1>2,3) x (2,3>1) y (3>3'); Independent: m (2, 3>2) x (3>3) n (1, 1>2,3) t (3>3') y (3, 3'>3) (295)	G	Person allomorphs are dependent on dependence or independence and participant scenario. Two sets of person markers (294, 295). "In Ayutla Mixe, dependent inflection is triggered when a non-argument appears before the verb. There-fore, whenever there is a locative adverb, a temporal or aspectual particle, the negative particle, or when there is an adverbial inter-rogative word before the verb, the verb is marked as being depend-ent. Conversely, a verb is marked as independent if only argumental constituents appear before the verb."(177)	(pos.)	Can be onset of the stressed syllable which is first in the word: ['mexp] m-ex-p 2s-see-INDEP 'you see' (296), also infixation/metathesis (with third person <y>, see [nya], morpheme gloss: y-na) (501); there are complex onsets (75).	yes	Person markers obligatory (291). They are mono-consonantal and do not change the syllable number (there is onset-cluster variabil-ity) (296)
-10	ADV	Non-nominal incorporation	nĕkoo 'only' ey 'good' tsuxk 'raw' (537, 538)	X		no	Syllabic, are not part of the stem sylla-ble	no	
-9	ADV	Motion cum purpose	ës (motion cum pur-pose) (376)	X		no	Syllabic, are not part of the stem sylla-ble	no	
-8	ADV	Directional and locative	e.g. kas 'downwards' yuk 'upwards' (377) në 'on' mu 'at' na 'circum-vention' (393)	X		no	Syllabic, are not part of the stem sylla-ble	no	
-7	V	Reflexive	nay (reflexive) (496)	X		no	Syllabic, are not part of the stem sylla-ble	no	
-6	V	Causative	ak (causative) (483)	X		no	Syllabic, are not part of the stem sylla-ble	no	
-5	V	Applicative	ta (instrument) (399)	X		no	Syllabic, are not part of the stem sylla-ble	no	
-4	V	Benefactive	më (benefactive) (512)	X		no	Syllabic, are not part of the stem sylla-ble	no	
-3	LEX	Incorporation	e.g. kafe 'coffee' nëj 'water' (531)	X		no	Syllabic, are not part of the stem sylla-ble	no	
-2	ADV	Part	e.g. ojt 'inside' këx 'out-side' (249)	X		no	Syllabic, are not part of the stem sylla-ble	no	
-1	LEX/ADV	Root (manner)	e.g. tem 'roll' (put) 'run' (602)	X		no	Not stressed, since not the last root of the verb.	no	

AYUTLA MIXE (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
0	LEX/ADV/A/M	Root, neutral completive, irrealis, independent, dependent, completive	See nucleus/coda alternation patterns (338)	L	Conjugational classes, phonologically defined but phonological shape does not determine class membership (e.g. CV'Vy can be either 6b or 7b) (338). Stem choice (e.g. completive) to express different TMA with position 7 (304, 305)	pos.	"In verbs (...) the last syllable of the verb stem is stressed. The verb stem is defined as what is left after the inflectional morphology is removed" Footnote: "The leftmost limit of the verb stem is defined by the last verb root in it." (80)	yes	
1	ADV	Phase roots	Incorporated phase roots, e.g. jëmpet 'return' (292) nëjk 'go' (292)	X		pos.	Never stressed, since these morphs are not part of the root.	no	
2	M	Desiderative	ä'ä(n) (desiderative) (364, 405). The desiderative is used for future reference (293)	U	No conditioning found (364, 405)	no	Never stressed, since it is not part of the root.	no	
3	ADV	Inverse	ë (inverse) (345)	X		no	Never stressed, since it is not part of the root.	no	
4	A	Perfective	n(e) (perfective) (308)	U	No conditioning found (308)	(pos.)	Perfective can be also only one consonant, as the coda of the (stressed) syllable of the stem, e.g.: y-kay-n (3sg-eat-PERF;DEP) (583)	no	
5	N	Plural	të/Ø (plural) (319	G	Allomorphs të or Ø dependent on mood/aspect (319).	no	Never stressed, since it is not part of the root.	no	
7	ADV/A/M	Independent, dependent, neutral, completive, irrealis	p (independent neutral intransitive) yp (independent neutral transitive) y (dependent neutral singular) t (dependent neutral plural) Ø (Independent Completive) (y) (Dependent Completive) p (Independent Irrealis) t (Dependent Irrealis) (304)	G	The interpretation and the expression of the (partly homophonous) markers is conditioned by another or two other categories (portmanteau markers) (304).	(pos.)	Coda of stressed syllable, like -t or -p (Table 7, p. 304)	yes	Obligatory (291), except in completive independent (Table 7, p. 304).

AZARI (GCI sample); Lee (1996)

Prominence: stress
"In most lexical or phonological words the primary stress falls on the final syllable, normally with a high pitch". "in some phonological words with enclitic suffixes the primary stress falls on the syllable immediately before the suffix (...) the present copular endings such as sAn sIz dIr; the second person imperative marked (y)In, the partcle da/də 'also'; the subordinator ki 'that', etc." (17)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
0.1	LEX	Root	Root, verb stem or noun can be incorporated with the help of an auxiliary in the following position (57-59)	X		no	Is never the last syllable of a word (cf. paradigms on p. 179)	yes	
0.2	LEX	Root (auxiliary)	Compound verbs like ib (drip) ye (eat); auxiliaries like ol (become) elə (make) et (show) sal (make/cause/drop) (58, 59) düş fall (59)	X		no	Is never the last syllable of a word (cf. paradigms on p. 179)	no	
1	ADV	Verbalizer	lA, A, ar, lAş, lAn (denominal verbalizers) (A)lt, Ar, lAş (deadjectival verbalizers) lAş (denumeral verbalizer) (55)	X		no	Is never the last syllable of a word (cf. paradigms on p. 179)	no	
3	V	Reflexive, reciprocal (order of valence suffixes (179)	y)(i)n/Il (reflexive) (57) (I)ş (reciprocal) (57)	X		no	Is never the last syllable of a word (cf. paradigms on p. 179)	no	
4	V	Causative, transitive	dIr (causative) t/It/Ir/Art (transitivizer) (55-56)	X		no	Is never the last syllable of a word (cf. paradigms on p. 179)	no	
5	V	Passive, reflexive, reciprocal	Il/(I)n (passive) (56)	X		no	Is never the last syllable of a word (cf. paradigms on p. 179)	no	
6	M/PO	Negative, impossibilitive	mA/mi/m (Negative) (y)AmmA/(y)Ammi/(y)Amm (Impossibilitive)	G	–mA appears in all verb forms but the present, aorist, future and optative forms; –mi appears in the future and optative forms, –m appears in the present and aorist, dropping the final –A before the following suffix vowel (52)	pos.	Can be stressed as it can be last syllable of the word, e.g. al-mi-yacax take-not-FUT 'he will not take' (52)	no	

AZARI (cont.)

Pos.	Cate-gories	Func-tions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
7	A	Perfect, pro-gres-sive	mIş (1.perfect) Ib (2/3.per-fect) (49) (y)Ir (Progressive) (51)	G	Perfect: Allomorph condi-tioning by person (49). Pro-gressive: no condtioning except phonological (50)	pos.	Can be stressed as it can be last syllable of the word, e.g. al-ma take-not 'Don't take!' (52)	no	
8	T/M	Opta-tive, neces-sitative, condi-tional, inten-tional, future	(y)A (optative) (48-49) mallI/məli (necessitative) (y)A bil (ability) sA (condi-tional) (51) (y)AcAX (inten-tional, future) (47-48; 51)	U	Optative: No conditioning except phonological (49). Conditional: no condition-ing except phonological (51)	pos.	Can be stressed as it can be last syllable of the word, e.g. ana-ymIş mother-in-ferential 'They say she is/was a mother.' (46)	no	
9	T/M	Present, inferen-tial, past, aorist	Ø (present) (46) y/ImIş (in-ferential) (46) dI (past) (46-47) (y)Ar (Aorist) (48)	U	Inferential: no conditioning except phonological (46) Past tense: No conditioning except phonological (46) Future: no conditioning ex-cept phonological (47-48); Aorist: No conditioning ex-cept phonological (48).	pos.	Can be stressed as it can be last syllable of the word, e.g. al-mi-yacax take-not-FUT 'he will not take' (52)	no	
10	P/N	Subject	Personal endings 1: m (1sg) n (2sg) Ø (3sg) X (1pl) z (2pl) lAr (2pl); Personal end-ings set 2: (y)Am (1sg) sAn (2sg) dI(r)/Ø (3sg) (y)IX (1pl) sIz (2pl) dI(r)/dI(r)lAr (3pl) (45) Personal endings set 3: (y)Im (1sg) Ø (2sg) (y)In (3.H) sIn (3sg) (y)AX (1pl) (y)In (2pl) sInlAr (3pl) (50)	G	Set 1 is used for predicate nominals or adjectives after past tense (dI) and condi-tional (sA). Set2 is used af-ter the suffixes of the pre-sent tense, the future tense (yAcaX), aorist (yAr) the perfect tense (mIş/yIb) and the optative yA and the ne-cessitive mali/məli (46-50) Set 3 is only used in the im-perative mood (50)	pos.	Can be stressed, except set 2 markers and (y)in in im-perative.	(yes)	"Agreement in person is always obliga-tory. Agreement in number requires close attention: it is obligatory, except for the third peson pural. It appears that the animacy hierarchy of subject noun phrase is affecting the agreement of plu-rality in the third person (122). Third person singular is zero, third person plu-ral can be zero.

BANGIME (LCI sample); Hantgan (2013) ("H"); Heath & Hantgan (2018) ("HH")

Most content words consist of (1) one syllable with two morae, (2) two syllables with either a light-light, light-heavy, or heavy-light combination, or (3) three syllables, all of which are light. There is a minimal word constraint of two morae, and a maximal word constraint of three syllables. Onsetless syllables are uncommon and codas do not occur word-finally. (H 76)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-4	P/N	Subject (clitic)	nɛ̀ (1pl) àà (2pl) nì (3pl) (HH 273)	U	Clitics seem to occur in all tenses/aspects (HH 292ff.)	no	2nd person plural can have a long mora (HH 273)	no	
-3	A/T/M/PO	Incompletive, completive, irrealis, negative, prohibitive, perfective (clitic)	daw/ dáà/dá/nà/ndà (incompletive), allomorphs after certain indexing prefixes (H308-312) kama/kóò (completive) (H 249) hà/há (irrealis) (H 252) bié (negative) (H 321) maa (prohibitive) (H 238) Ø (perfective) (H 238)	G	The completive allomorphs are conditioned by the existence of a following object noun phrase/preceding indexing clitic (HH 282). However, there seems to be no hard conditioning for kóo/kama (H246).	pos.	Long mora for example kóò perfective (HH 272)	no	
-2	NP	Object (noun phrase), direct/reflexive object	Object can intervene in some word orders (H 260). Object/reflexive pronouns: mí (1sgO) á miì (3, 1pl reflexive) aà (2pl) (438-439)	G	In imperfective constructions, it occurs between P-3 and P-1, in future construction, the direct object occurs between P-4 and P-3 (HH 342).	pos.	Objects (phrases) include nouns which can have different moras.	no	
-1	P/V	Transitive, Subject (clitic)	m (transitive) (H 233) ŋ/n (transitive, non-second person) (HH 278)	G	Appearance is governed by grammatical category: "In an incompletive clause, the subject is marked by a phrase-initial nasal for non-second persons, and the transitive nasal marker precedes a transitive verb." (240) (example 291 on H p. 240 shows that the transitive marker in this position is analyzed as a first person marker).	no	All personal pronouns here have a short mora (HH 273)	no	Only in incompletive clauses (240)

BANGIME (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
0	LEX/V/A/M/P/N	Root, incompletive, completive, perfect, perfective, subject, causative, deontic	Root modification (incompletive), root modification (completive) root modification (perfect), root modification (perfective) (H 198). Person and number subject are expressed with tones (H 278). Deontic/Modal tones (HH 295) ndV (Causative, replaces root final vowel) (HH 211) (See examples for root alternation e.g. H 207 and 2012). "There are 496 verbs in the corpus. Verb roots, like noun roots, do not surface without affixation and tonal overlays. Most verb roots surface with either an additional vowel to meet the minimal word requirement of two morae, or with morphological marking in the form of inflectional suffixes." (198)	L	Summary of verb classes (214) with several endings depending on verb class (e.g. -da can be completive in 3rd verb class and incompletive in 4th class). Also person marking is usually dependent on the tone of the verb, but some verbs (H 284) have a different tonal pattern for verbs (285).	pos.	Verbal roots are minimally bimoraic, of which the last mora can be a suffix or a root element. They can be the most trimoraic, in which case either the first or the second syllable are heavy/light, or all the syllables are light (see H 84 for minimal word constraint).	yes	with "yes" a root with maximally three moras is meant. These can be either all part of the root (HH 235) or root plus a suffix (e.g. tuu tura HH251). The suffixes will be treated as stem alternation since they do not expone more than two or three moras.
1.1	P	Reflexive	Object/reflexive pronouns in perfective negative: mí (1sgO) á miì (3, 1pl reflexive) aà (2pl) (HH 439-438)	G	Object/reflexive pronouns only in perfective. (HH 438-439)	pos.	Some of the object pronouns have a long mora (3rd, 2pl). (HH 438-439)	no	
1.2	P	Subject (clitic)	ŋ (non-second person) see person suffix conjugation in (H283)	G	Same as CL-1. Appears with perfective kéè. (H210)	no	All personal pronouns here have a short mora (HH 278)	no	
2	A/T	Future, perfect, resultative (clitic)	naw (future) ke (perfect) (H 210) wàj (perfect stative/resultative) (H 210). For positioning after the person clitic see HH 277.	U	These suffixes appear together with some of the position -2 clitics but only the -2 clitics' meaning is dependent on them, not the other way round.	no	All forms seem to be monomoraic. The older version (H 276) writes kee as bimoraic but in the newer version (HH 271) it is found as monomoraic.	no	

BARDI (GCI sample); Bowern (2004)

Prominence: stress

"Stress is usually on the initial vowel of the word" (87)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-6	ADV/T/M/P/N/T	Subject	nga (1) mi (2.present/past) a (2.bivalent.future/irrealis/imperative) nga (2.monovalent.future/irrealis/imperative) i (3.past/present) oo (3.future/irrealis) a (1+2; 1.aug) goo (2.aug) (180-181) ma (gerund/infinitives) (205)	G	Person allomorph dependent on tense/mood/transitivity (see tables on pp. 180-181)	yes	The initial vowel of the verb is always a subject marker, since subject markers are obligatory (100) and syllabic (180)	yes	"All inflecting verbs in Bardi show a person prefix, and a tense prefix" (100)
-5	V	Transitivity (clitic)	Ø (intransitive) n (transitive) (180-181)	X	Occurs if there is no augmented marker following (p. 347)	(no)	The initial vowel is never part of this position (The initial vowel is always part of pronominal elements in P-6). n– does not occur word-initially.	no	
-4	T/M	Past, present, future,iIrrealis	ng (past) Ø (present) (ng)g (future) l (irrealis) (180-181)	U	No conditioning found except phonological (104, 211-216, 346, 347)	(no)	monoconsonantal, always coda, never onset of the first syllable(346-347)	(yes)	"All inflecting verbs in Bardi show a person prefix, and a tense prefix" (100). Only present tense is zero (346-347)
-3	N	Subject	Ø (minimal) rr (augmented) (180-181)	U	No conditioning found (180)	(no)	monoconsonantal, never initial segment of a word (347)	no	
-2	V	Transitivity (clitic)	Ø (intransitive) n (transitive) (180-181) m(a) (reflexive/reciprocal) (230f.)	X	Occurs if there is an augmented marker preceding (p. 347)	no	The initial vowel is never part of this position (The initial vowel is always part of pronominal elements in P-6)	no	
-1	ADV	Pluractionality	full/partial reduplication (dependent on the syllable count of the root) (iterative, distributive, instensive, plural participants) (148-154).	X		no	The initial vowel is never part of this position (The initial vowel is always part of pronominal elements in P-6)	no	
0	LEX	Root	Can be: light, simplex, derived, historically complex, incorporated (158)	X		no	The initial vowel is never part of this position (The initial vowel is always part of pronominal elements in P-6)	yes	

BARDI (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
1	V	Applicative	ng(a) (applicative) (238-239)	X		no	The initial vowel is never part of this position (The initial vowel is always part of pronominal elements in P-6)	no	
2	V	Reflexive	inyji (reflexive) (231)	X		no	The initial vowel is never part of this position (The initial vowel is always part of pronominal elements in P-6)	no	
3	A/T	Continuative, remote past	n (continuative) (221) na (remote past) (222)	U	Continuative: no conditioning found (221). Remote past: no conditioning found (222-223)	no	The initial vowel is never part of this position (The initial vowel is always part of pronominal elements in P-6)	no	Zero-marking of T/A is possible (225-226)
5	V	Applicative	ng(a) (applicative) (239-242)	X		no	The initial vowel is never part of this position.	no	
6	A/T	Recent completed past, middle perfect, future	gal (recent completed past) (217) ij (middle perfect) (223) a (future) (101)	U	Recent past: No conditioning found (217). Middle perfect: no conditioning found (223). Future: no conditioning found (223-224)	no	The initial vowel is never part of this position (The initial vowel is always part of pronominal elements in P-6)	no	Zero-marking of T/A is possible (225-226)
7	A	Simultaneous, subordinator	j (simultaneous; subordinator) (224)	X		(no)	The initial vowel is never part of this position (The initial vowel is always part of pronominal elements in P-6)	no	

BARDI (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
8	ADV	Linker, relator, contrastive, resumptive	b(a) (linker) (j)amb (relator) min (contrastive) gid (resumptive) (105)	X		no	The initial vowel is never part of this position (The initial vowel is always part of pronominal elements in P-6)	no	
9	ADV	Topic	jarr/ji(rr) (topic) (prefixed to either direct or oblique object speech act participants) (191, 203)	U	No conditioning found except phonological (191, 203). Usually attached to prefixes that are not zero (i.e. not already topical) (194)	no	The initial vowel is never part of this position (The initial vowel is always part of pronominal elements in P-6)	no	
10	P/N	Oblique	jan (1min) jiy (2min) jin (3min) jow (1+2min) jard (1aug) joogarra (2aug) jirr (3aug) (102)	U	No conditioning found (201-204). The two sets on p. 203 are analyzable as topic + oblique. The reason why the whole table is not filled out with these forms is probably due to their low frequency; "The only attested forms are 1MIN and 3MIN)" (203)	no	The initial vowel is never part of this position (The initial vowel is always part of pronominal elements in P-6)	no	
11	P/N	Possessor	(j)an (1min) (j)iy 2min (j)in (3min) (j)ow (1+2min) (j)ard (1aug) (j)ard (1aug) (j)oogarra (2aug) (j)irr (3aug) (102). These forms are used for possessor raising (204).	U	No conditionig found (204-205).	no	The initial vowel is never part of this position (The initial vowel is always part of pronominal elements in P-6)	no	
12	P/N	Direct object	ngay (1min) rri (2min) Ø (3min) yow (1+2min) moordoo (1aug) gorr (2aug) irr (3aug) (102).	U	No conditioning found (190-197). The two set forms on p. 191 are analyzable as topic+object marking.	no	The initial vowel is never part of this position (The initial vowel is always part of pronominal elements in P-6)	no	
13	ADV	Quantifier	nid (many) (b)al (indefinite) (tend to refer to subject or object of the clause) (106).	X		no	The initial vowel is never part of this position (The initial vowel is always part of pronominal elements in P-6)	no	

BETOI (LCI sample); Zamponi (2003)

Prominence: stress
"Stress is limited to occurring on one of the last two syllables of the word, falling more frequently on the final one, and that in long, polymorphemic words, other syllables probably take a weaker level of stress." (11)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-2	T	Past, future	fa (FUT) ma (PAST) (25)	U	No conditioning found (25)	no	Does not bear stress since it precedes the stem.	no	
-1	P	Subject	r (1) j (2) Ø (3) (29) (Table 12, p. 22)	G	Active and stative verbs with vowel initial stem in P0. Not active verbs that have both stems in P0 and P2, but stative verbs.	no	Does not bear stress since it precedes the stem.	no	Only present in active and stative verbs with vowel-initial stem. (22)
0	LEX	Root	First part of theme. Vowel initial and consonant initial trigger different agreement markers (22).	X		pos.	Some verbs have secondary stress on the stem (e.g. rááquirra-bica-rrú, p. 11)	yes	This part of the verbal complex is always filled (22)
1	P/N	Subject	rr(a) (1sg) j(a) (2sg) Ø (3)	G	Active verbs for consonant initial stems in P0 (Table 13, p. 22). Active verbs with stem in P2 (Table 14, p. 22). Stative verbs if there is a stem in P2. (30)	no	Table 13 on p. 22 does not reveal that they could be stressed (following affixes in the non-kernel verb can have stress)	no	Only in some verbs (22)
2	LEX	Root	Second part of active verbs (22)	X		pos.	Can be stressed, see fa-rr-inefá FUT-1-take.care.of "I will take care of" (40)	no	

BETOI (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
3	P/N	Subject (vowel-initial active verb subject markers without bipartite theme)	rra (1sg) ja (2sg) Ø (3sg) mai ~nutó (1pl) jui ~jai (2pl) bi (3pl) (Table 12 and 13) (22)	G	Active verbs, with vowel initial stem in position 0	pos.	nutó can bear stress (22).	no	Only in active verbs with vowel initial stem (22)
4	PO	Negation	omé (NEG) (34)	U	No conditioning found (34)	yes	Always stressed either on the /e/ or /o/ like falabómelú (35).	no	
5	A/M	Indicative, conditional, purposive/optative, prohibitive, imperative (zero), imperfective.	cá (IND) idaódda ~ idódda (COND) dianú (PURP) ometú (PROHIB) Ø (IMP) ida (IMPF) (21)	G	Purpositive/optative have several allomorphs that change depending on person (26).	pos.	Allomorphs can bear stress (see paradigms on p. 21)	no	
6	N/G	Number of subject, singular/masculine/feminine/neuter, plural (29)	oi (sg.M), ó (sg.F), ajoi/aje (sg.N), olá/oladá (pl) (29)	G	Only in statives. "Due to the paucity of data, the linguistic contexts of the variation -ajoi ~aje 'SG:N' and olá ~oladá 'PL' cannot be recovered". That means, it canot be known whether the alternation is lexically conditioned (29).	pos.	Allomorphs can bear stress (see paradigms on p. 29)	no	Only in stative verbs (26)
7	P/N	Subject, (in)direct object	rrú (1sg) nutó (1pl) jú (2) Ø (3) (stative or active subject) (29)	L	Subject for some stative and some active verb (29): object or indirect object for active verbs (p. 27). The occurrence of these morphemes is lexically determined, see verbs that have it (30).	yes	Affixes are always stressed, as they are the last affixes and stress is always on the final syllable (see paradigms on p. 29)	no	In some statives (29), active verbs (21). However, not all statives have this position, and it's not predictable why (p. 30 for examples of stative verbs)

BURUSHASKI (LCI sample); Berger (1998)

Prominence: stress
First or second syllable is stressed. Details in the explanation for prominence assignment.

Pos.	Categories	Functions	Morphs/Features	Cond.	Examnple Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-4	PO	Negation	á, óo (NEG) (106)	L	oó can be fusion of a- and u- or a- and d- (106). However, the morph does not fuse with the stem.	pos.	Negative morpheme can be additionally stressed, like oó-dimélyaljan 'we will not listen.' (107)	no	
-3	ADV/V	Transitivity	d(u/i) (intransitive) n(u) (infinitive) (107)	X		pos.	Can be either first or second syllable.	no	
-2	P/N/G	Subject	a (1sg) gu (2sg) i (3sg) mu (3F) mi (1pl) ma (2pl) u (3pl.hum) i (3pl.nonhum) (111)	L	Pronominal Subject prefixes for some intransitive verbs. The semantics of the specific verb class that use it is only partially predictable (117)	pos.	Accent on second syllable, some prefixes here can be accented (105)	no	
-1	V	Transitivity	s (TR) (125)	X		(no)	Monoconsonantal, always appears with prefixes (125), therefore never word-initial (125)	no	
0	LEX/T/M	Root, tense, mood	stem (125), suppletion for number, tense (present, past, plusquamperfect), infinitive (128), mood (subjunctive, indicative) (129) for verbs: eat/slaughter, come, go	X		pos.	Two-syllabic stems are usually stressed on the second syllable of the stem. Monosyllabic stems bear stress on the stem (127) - when they occur without prefixes - or on the peg element in P3. (130)	(yes)	"In manchen Fällen kann der Verbalstamm ganz verschwinden." (Berger 1998: 128)

BURUSHASKI (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
1	N	Plural	ya/za (pl) i (peg element) (130)	L	Additional plural marking for certain (12) verbs (Otherwise Ø-Allomorph). Shape allomorphs phonologically conditioned. (130)	pos.	After monosyllabic stems without pronominal prefix but plural suffix in P1, this plural morpheme is stressed (130).	no	
2	T	Present	ch/sh/j/c/y (present) (130)	G	Different allomorphs according to different ending consonants of the stem, also is used for future, present, imperfective and conditional (130)	(no)	monoconsonantal, never appears without a suffix (130).	no	
3	P	Subject, peg element	(y)a (1) a/u/i (peg element) (132)	G	Peg vowel element dependent on following (P4) categories: participle, optative, finiteness and dependent on the last consonant of the stem (132).	pos.	Peg element can be stressed man-ú-ma 'You became' (132)	yes	Lorimer 1935: 264, even in intransitive with prefixes this vowel is obligatory.
4	T/M	Participle, infinitive, optative	m (participle, TMA) n (infinitive) (a)as (infinitive) sh/r/l/n (optative) aa (optative) (133ff.)	U	m-suffix used for future and conditional and optative, but there is no allomorph for m. Conditional, optative and future are semantically closely related, therefore this is not seen as grammatically conditioned (133).	pos.	Optative aa bears additional accent (136)	no	
5	A/T/M/P/N/G	Oblique	b(a) (auxiliary) a (2sg) i (3.hum) o (3sg.F) ie/ila (3pl.nonhum) an (hum.pl) (136 ff.)	G	Personal endings dependent on optative, conative, imperative, auxiliary, negative, future (136ff)	no		no	First person is expressed in P 3.
6	ADV	Interrogation, nominalization	a (interrogative), case endings for nominalized verbs (140 ff.)	X		no		no	

DUMI (LCI sample); Van Driem (1993)

Prominence: stress
"Unless indicated otherwise, verbs and deverbatives are always stressed on the root. (...) Affixes, whether flexional or derivational, are never stressed." (58)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-2	P/N/V	Subject, marked scenario (inverse)	ham (3pl intransitive or reflexive) a (marked scenario) (121)	U	No conditioning found (122)	no	Never stressed, since not root morphemes.	no	
-1	T/PO	Negative, past	mə (negative past) (121)	G	Does not occur when a occurs in P-2. (124)	no	Never stressed, since not root morphemes.	no	
0	LEX/ADV/T/P/N/	Root, subject, pluractionality	Pluractionaliy through reduplication (129)	L	Stem alternations according to different lexical conjugation classes (5 intransitive, 7 transitive conjugations). "Once the conjugation of any given verb has been specified, it is predictable which stem will occur i a given inflected form" (91). Stems alternate for their stem vowel, final consonant or both (92). The alternation is different in different conjugations, affecting the syncretism in number, person, and tense (see tables e.g., p. 96). For example, the third intransitive conjugation marks 1st person singular, dual, 2nd person and 3rd person with stem1, 1st person plural with stem 2. Verbs of the 5th intransitive conjugation have the same distinction, except that within first person plural there are different stems for non-past and past (stem 3 is used for 1pl Past). (96)	yes	Always stressed, since root morpheme.	yes	
1	V	Reflexive	nsi (reflexive) (121)	X	Phonologically conditioned allomorphs (125)	no	Never stressed, since not root morphemes.	no	

DUMI (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
2	P/N	Subject, object	k (1pl) n/Ø (1sg > 2) ŋ (1sg) (121)	L	After vowel-final stems of verbs belonging to transitive conjugations 1, 5,6b and 7, the 1sg>2 portemanteau occurs in its neutral form /n/, After vowel-final stems of verbs belonging to transitive conjugations 4 and 6a, the 1sg>2 morpheme <-n> is realized as /n/ in Is-»2s and ls-»2p forms but as zero in 1sg>2du forms (132) ŋ (1sg>3) "occurs in all open stem verbs except open stem vi-1 verbs" (134) n, ŋ and k also occur as infixes in aspectivized compounds (134). "The first 1s suffix <-ŋ> occurs in all open stem verbs except open stem vi-1 verbs." (134)	(no)	Usually coda of the stem (131), but never beginning or final segment of a word, since other affixes must follow (see paradigms on p. 151-167).	no	
3	V/P/N/C	Reflexive, subject, object	si (reflexive) n (1sg>2) si (2/3du) i (inclusive) ɨ (exclusive) (121)	U	Appears when there is no t (non-past) ('copy slot') (121) No conditioning except phonological (146-148)	no	Never stressed, since not root morphemes.	no	
4	T	Past, non-past	t (non-past) Ø (past) (121)	U	No conditioning found (135-140)	(no)	Never stressed, since not root morphemes. -t never occurs at the end of the word (see paradigms on p. 151-167).	no	
5	T/P/N/C	Subject, object	ə (1sg) u (1sg>3.PAST) i (inclusive) ɨ (exclusive) a/Ø/ɨ (2/3.subject) ɨ/Ø (3sg.patient.PAST) (122, 143, 144)	U	No conditioning found except phonological (140-144)	no	Never stressed, since not root morphemes.	no	
6	T/P/N	Subject, object	i (du) a/Ø (2/3du.nonPAST) ini (2/3pl) (122)	U	No conditioning found except phonological (145-146)	no	Never stressed, since not root morphemes.	no	
7	PO	Negation	nə (NEG)	U	No conditioning found except phonological (149-	no	Never stressed, since not root morphemes.	no	

ESTONIAN (LCI sample); Blevins (2007) ("B"); Harms (1962) ("H")

Prominence: stress
"Phonemic stress (...) is manifested within the vocable-initial syllable" (H 11)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
0	LEX/ADV/T/M/P	Root, infinitive, present, supine, impersonal, jussive, indicative, past, conditional	Gemination of last consonant depending on infinitive, present, supine or impersonal (B 254) Vowel alteration in conjugation class (<oe> in present, impersonal and <uge> in infinitive and supine) (B 255)	L	Whether a consonant is geminated or not depends on the conjugation class and inflectional category (B 254)	yes	"Phonemic stress (...) is manifested within the vocable-initial syllable" (H 11)	yes	
1	ADV/A/T/M/PO	Person, number, present, conditional, past, imperfective, infinitive, supine	Present/Indicative: n (1sg) d (2sg) b (3sg) me (1pl) te (2pl) vad (3pl); Present Conditional: ksin (1sg) ksid (2sg) ks (3sg) ksime (1pl) ksite (2pl) ksid (3pl); Past/Imperfect: sin (1sg) sid (2sg) s (3sg) sime (1pl) site (2pl) sid (3pl) (252); da/ta/a (Infinitive) ge/ke (2pl Imperative) gu/ku (Jussive) nud (Past participle) ma (supine) v (present participle) vat (evidential) da/ta (impersonal negative indicative) dakse/takse/akse (impersonal present) di/ti (impersonal past) dud/tud (impersonal past participle) gu/tagu/dagu impersonal imperative) (B 252)	L	Third person indicative, conditional and past allomorphs except impersonals dependent on the grammatical category (B 252). Other allomorphs depend on the conjugation class, e.g. jussive -gu in õppigu (study) and lugegu (read) but -ku in hüpaku (jump) and lennaku (fly) (B 257)	no	Never the first syllable of the word, since the stem is obligatory.	yes	There are no zero paradigms (252)

FRENCH (STANDARD) (LCI sample); Batchelor and Chebli-Saadi (2011); Ashby (1977) for positions;

Prominence: Stress
"Slight stress on the last articulated syllable of groups of words" (B 45)

Positions: Finite Verb (Ashby 1977: 16) (without imperative, infinitive, inversion). Inversion is a more marked language speech that is used in literary style (B 496). Without these sentence types, the proclitics can be counted as prefixes

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-11	P/N/G	Subject pronouns	ʒə (1sg) ty (2sg) il (3sg) sa/se (3sg.def) nu(z) (1pl) õ (1pl/3.indefinite) (Subject) vu(z) (2pl) i(z) (3pl) (A 16)	U		no		no	Third person pronoun can be elided but this is not very common (443). Usually people repeat the pronoun even when a noun phrase precedes (442). Third person pronouns are, however, not obligatorily present with noun phrases.
-10	PO	Negation	nə (Negative) (A 16)	U		no		no	
-9	T	Future	vɛ (FUT.1sg) va (FUT.2/3sg) alõ (FUT.1pl) ale (FUT.2pl) võ (FUT.3pl) see examples 1-3 on A 18) (paradigm B 741; usage B 258)	G	Different allomorphs for different persons.	no	Different allomorphs for different persons.	no	
-8	ADV/PO	Negation/ intensity (clitic)	gɛr (B 528; Intensity) jamɛ (never) nəply (any more) pwã/pa (negative) rjɛ̃ (nothing) pɛrsɔn (noone) (different negatives) (A 16) (usage B 637ff.)	G	When there is the auxiliary in P-9.	no	When there is the auxiliary in P-9.	no	

FRENCH (STANDARD) (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-7	P/N/G	Direct/ indirect object	mə(1sg) tə (2sg) sə (3sg) nu (1pl) vu (2pl) (A 16) (B 440)	U		no		no	
-6	P/N/G	Direct object	lə (3sg.M) la (3sg.F) le (3pl) (A 16) (B 440)	U		no		no	
-5	P/N/G	Indirect object	lɥi (3sg) lœr (3pl) (A 16) (B 440)	U		no		no	
-4	P/N	Location	i (location) (A 16) (B 440)	X		no		no	
-3	ADV	Partitive object	ɛ̃ (some) (A16) (B 440)	U		no		no	
-2	A/T/M/P/N	Perfect, mood, subject (auxiliary)	Perfect auxiliary: intransitive (être): sɥi (1sg) e (2/3sg) səm (1pl) ɛt (2pl) sõ (3pl); transitive (avoir): ɛ (1sg) a (2/3sg) avõ (1pl) ave (2pl) õ (3pl) etre/avoir conjugated, depending on state or action; conjugated for the following tenses: pluperfect, past anterior, future in the past, conditional in the past (B 720). Modal auxiliaries (1sg): dwa (must) pwa (can) vwa (want); inflected for tense, mood and person, e.g. pudrɛ (could) vudrɛ (would). (B 339) (A 16)	G	Auxiliary depends on transitivity, mood, tense and person. "avoir" conjugated with transitives (B 197) and "être" with intransitives: "intransitive verbs associated with movement or change in a particular state and (2) pronominal or reflexive verbs." (B 200)	no		no	

FRENCH (STANDARD) (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-1	PO/ADV	Negation, adverbs of quantity or intensity (clitic)	gɛr (B 528; intensity), jamɛ (never), nəply, (any more) pwã/pa (negative) rjɛ̃ (nothing) pɛrsɔn (noone) (different negatives) prɛsk (almost) when auxiliary present in P -2 (A 16) etc. (use B 528 637ff.)	G	When auxiliary is present in P-2.	no		no	
0	PO/A/T/M/P/N	Root, tense, mood, aspect, subject/object	Root, vowel alternation and/or dropping/retaining last consonant of some verbs in plural forms, like sɛ (know.1sg.IND) savõ (know.1sg.IND) (B 745) pø (can.1sg.IND) puvõ (can.1pl.IND) (B 743) mœr (die.1sg.IND) murõ (die.1pl.IND) (B 740)	L	Dependent on verb/person/mode/tense/participle (see conjugation tables B 723ff.)	pos.	Can bear stress if no suffixes or clitics follow, e.g. [il'pøv] 'they can'	yes	
1	A/T/M/P/N	Tense, mood, aspect, subject/object	Mode and person suffixes, irregular. (i)õ (1pl) (i)e (2pl) i (1/2/3sg); tenses/moods (1sg): i (preterit) rɛ (future/conditional), see for example paradigms in (B 720). te/y/i Participle suffix to mark past tense (B 720-722)	L	Choice of this suffix depends on the verb (conjugation) but also on the interplay of mood, tense, person, number (see tables 719ff.).	pos.	Can bear stress if no suffixes or clitics follow, e.g. [nupuv'õ] 'we can'	no	Zero in singular (all persons) and plural (third person) indicative.
2	ADV	Adverbs of quantity/intensity	absolymã (totally), dõk (then) mɛm (only) prɛsk (almost), pur ɛ̃si dir (to say so) (A 16) etc. (B 528)	X		pos.	Can bear stress if no clitics follow.	no	
3	ADV/PO	Negation, intensity (clitic)	gɛr (B 528; Intensity), jamɛ (never), nəply, (any more) rjɛ̃ (nothing) pwã (Intensive negation) persɔn (noone) (different negatives) (A 16) (use B 637ff.)	G	When there is no auxiliary in P-2.	yes	Stressed, as being the last word in the group of words.	no	

HATAM (UCI sample); Reesink (1999)

Prominence: stress
"Accent in Hatam cannot be stated in terms of word pattern. It is related to a whole utterance or phrasal part thereof. (32). "Content words, like monosyllabic verbs or nouns, do not necessarily attract stress" (33) "With bisyllabic words, the first syllable is never stressed " (32), except for monosyllabic lexicals with -ni and -a suffixes following, the first syllable is stressed (32)

Pos.	Catego-ries	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-4	ADV	Derivation (clitic)	gi (nominalization) (46) di (relativizer) (47)	X		no	"Neither [gi or di] are ever tressed" (43)	no	
-3	ADV/T	Posterior (future), purpositive (resulta-tive) (clitic)	mi (posterior) (55) bi (pur-positive/resultative) (47)	U	No conditioning found (55)	no	Never stressed as it rep-resents a weak syllable (31)	no	
-2	P/N	Subject	di (1sg) a (2sg) Ø/ni (3sg) si (1du) i(g) (1inc) ni (1exc) ji (2pl) (i)g (3pl) (40-41)	U	No conditioning found (40-41). There is no mentioning on how the alternation Ø/ni in 3rd person might be condi-tioned, very probably phono-logical (see p. 23)	no	Never stressed as the pronouns are all weak sylables (31)	(yes)	"Verbal predi-cates are al-ways marked for person-number of the subject." (87)
-1	ADV	Instrument	bi (instrument) (54)	X		no	"[bi] behaves phoneti-cally as the subject pre-fixes. That means it is never stressed" (54)	no	
0	LEX	Root		X		pos.	"Content words, like mono-syllabic verbs or nouns, do not necessarily attract stress" (33)	yes	

INGUSH (LCI sample); Nichols (2011b)

Prominence: stress
Stress is marked almost always on the first syllable. (96)

Pos.	Catego-ries	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explana-tion Obl.
-4	ADV	Deictic	dwa (away from speaker) hwa (toward speaker) hwal (up) wa (down) (320)	X		yes	Stress is marked al-most always on the first syllable. Since these morphs are sin-gle words, they are always stressed (96)	no	
-3	LEX/ADV	Local prefix, lexical prefix, lexically in-corporated element, reduplicated root, heavy (lexical) piece of compound verb, syntactically incor-porated element	aara (out) juxa (back; again, over, re-) uragh (up, upwards) laqie (above, up) loxie (below, down) aarq'al (face up) (337) hwal (up) (335) t'y (on) (336)	X		no	Deictics are atonic (97)	no	
-2	ADV	Negative (can occur in P-1, 639), syntac-tically incorporated element	cy (negation) (343, 639) my (negative impera-tive) (341) myshta (how) mycha (where) mel (how much) maca (when) (341)	G	Negative morpheme only occurs in non-finite con-structions (infinitive, converbs) (631), otherwise as a suffix (101)	no	"The proclitics my= (…) and cy= do not have stress (5) but have high tone (737).	no	

INGUSH (cont.)

Pos.	Catego-ries	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Ex-pla-na-tion Obl.
-1	ADV	Serializing particle (can occur in P-2, 639)	a (639)	X		no	Serializing particles are attached to the preceding word (negation or incorporation) and therefore do not bear stress	no	
0.1	LEX/G	Root, absolutive	Gender morphemes: j v b d (133)	G	Nichols analyses the agreement marker as part of the stem, since it replaces the onset of the verb (303). Onset of stem except v-, d-, j-, b- does not inflect for these four genders. Therefore, whether there is agreement or not on a verb is not lexically, but phonologically conditioned. The morphs are conditioned by the specific role and gender of the referent, which makes them grammatically conditioned. The gender assignment of referents is not straightforward, except for v (masculine) and j (feminine) gender (132).	(pos.)	Consonantal (can be first consonant of a syllable that bears stress at the beginning of the word)	yes	
0.2	LEX/A/T	Root, present, simultaneous, infinitive, verbal noun, witnessed past, anterior converb	Different vowel alternation patterns (220)	L	Regular verbs fall into 16 ablaut classes which are lexically conditioned (219)	pos.	Bare verb stem can be the first syllable of a word.	yes	

INGUSH (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
1	V/ADV	Causative, Inceptive	Gender-u (direct causative, make) (i)t (leave, indirect causative) gender-eit (make+leave, double causative) lu (give, inceptive) (443f.) gender-oliit (indirect causative) gender-eit (double causative) (318)	X		pos.	Is never first syllable of the verb, since roots are minimally monosyllabic (222); however, causative -(i)it attracts stress. (96)	no	
2	T/M/A	Present, imperfect, future, simultaneous, sequential, infinitive, past, sequential, nonwitnessed past (auxiliary)	Present stem: Ø/a/V (present, last two are irregular) (a)r(a) (imperfect) (a)(r)g (future) (a)zh (simultaneous), ie/j (sequential) a; Infinitive stem: (infinitive/imperative); past stem: (a)ra (witnessed past), aa/(a)ra/Ca/na (anterior), ie (sequential) (222) ad/aa+d (Nonwitnessed past + gender) (101)	L	Suffixes' interpretation depends on the stem type (222). Also, some of the suffixes (present, imperfect, witness past, anterior converb) are restricted to some specific set of verbs (see restrictions p. 222, e.g. for present tense, the final vowel of the stem is sometimes -a, but not always, p. 221)	pos.	In witnessed (227) and non-witnessed past (pluperfect construction) high tone (rise-fall) get reanalyzed by some speakers as stress (96, 97)	(yes)	obligatory (218), although present tense has a zero allomorph, whereas other morphs are syllabic.
3	A/T/M/PO	Future, conditional, past inferential, future progressive, negative phasal verbs (auxiliary)	Inflected TMA auxiliary, Gender+y (future) (228) Gender+ar (finite conditional past) (231) xudd[Gender]ar (past inferential) (232) [Gender]ar (future progressive) (244) -ac (negative present) andz (negative witnessed past) al (imperative) (101) (phasal verbs (530) not included)	G	Interpretation depends on stem type and combination. e.g. Future progressive: 'simultaneous converb plus future tense of delimited 'be' (244).	no	Is never first syllable of the verb.	no	

K'ICHEE' (GCI SAMPLE); MONDLOCH (2013)

Prominence: stress
"In K'ichee' words are generally stressed on the last syllable (e.g., tinamít -- "town"; kixqatzukuúj -- "we look for you"). However, small particle-like words are generally not stressed unless they are the last word in an utterance (e.g., with the adverbial particle, na -- "later": kimb¢eé na chwé¢q; utterance final -- kimbeé ná). If a word ends in a short vowel, then the stress is on the penultimate syllable of that word (e.g., kojeewára jee la' "We'll go sleep over there." , chúwa ja) "in front of the house". " (10)

Pos.	Catego-ries	Func-tions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-3	A/M/PO	Incom-plete, complete	Ø (perfect) (141) k (incomplete) k/ch (imperative) (30) x (completed) (34) ch (imperative) ch (imperative transitive) m (negative imperative) (136)	U	No conditioning except phonological (136)	no	Is never the last syllable of a word, therefore never receives stress.	(yes)	Aspect marking is always present, with an exception of the zero-marked perfective (141)
-2	P/N	Absolu-tive	in (1sg) at (2sg) la (2sg.formal) a/Ø (3sg, 2 formal) oj (1pl) ix (2pl) alaq (2pl.formal) ee (3pl) (30)	U	No conditioning except phonological (30)	no	Is never the last syllable of a word, therefore never receives stress.	yes	All verbs have an absolutive argument. Third person is zero. (30–97)
-1	P/N	Ergative	in/w/inw (1sg) aa(w) (2sg) uu/r (3sg) q(a) (1pl) ii(w) (2pl) ki() (3pl) (53, 62)	U	No conditioning except phonological (53, 62)	no	Is never the last syllable of a word, therefore never receives stress.	no	Only in transitive clauses (52–97)

K’ICHEE’ (cont.)

Pos.	Catego-ries	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
0	LEX	Root		X		pos.	Can be the last syllable of the word, therefore, it can be stressed. The root is also stressed when a monovocalic suffix follows (kojeewára jee la’ “We’ll go sleep over there.” (war = sleep)) (10)	yes	
1	ADV/V	Intransi-tive, tran-sitive	x (passive voice from derived transitives) (69) j (transitive) (54) u/o (transitive final for monosyllabic roots) (54, 108) taj (com-pleted passive) (74) tal (positional aspect, stative) (75) n (Antipassive) (79) o'w (agent focus antipassive) with monosyllabic stems (132) i' (positional intransitive) (156) l (po-sitional transitive) (156)	X		pos.	Can be the last syllable of the word, therefore, it can be stressed.	no	
2	A/M	Impera-tive, indic-ative, per-fect, re-sultative	oq (imperative in final position, intransi-tive) a (imperative non final) (41) oq/qa (imperative passive for monosyllabic stems) (122) a'/o'/u' (imperative for transi-tive monosyllabics) (115) Ø/(i)k (indicative final position for intransitives, passives, an-tipassives) (31) V(V)m (perfect) (141) (i)naq (resultative/perfect passive) (147)	G	Only imperatives have allo-morphs depending on va-lency(41, 115, 122). Other morphs have phonological condi-tioning (sentence final/non-final position)	yes	All suffixes do not end with a vowel and are at the end of the word (10).	(yes)	Final element not present if the verb is not in final position
3	ADV/P/N	Subject impera-tive, loca-tion (clitic)	la (2sg.formal) alaq (2pl.formal) (30, 53), b'a(') (exclamatory) (41), wi (to) (49)	U	No conditioning except phono-logical (sentence final/ non-final) (30, 53, 41, 49)	pos.	Only when particles are at sentence-final position (10).	no	

KET (LCI sample); Vajda ("V") 2001; 2003, 2004; Werner ("W") (2002)

Prominence: pitch contour
Verbs have a characteristic rising-falling pitch contour and can repeat this contour in multisyllabic verbs (V 2003: 22). The rising and falling tones both have the acoustic impression of stress (V 2003: 16). For all inflectional positions in the template: "Each stem lexically selects a subset of these [inflection] slots filling them with the syntactically appropriate subject or object markers." (V 2004: 45)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-8	P/V	Subject, involuntary causative	d(i) (1) k(u) (2) d(u) (3.M/3pl) da (3sg.F/N/Involuntary causative) (V 2004: 48,68)	L	See description above.	pos.	Can bear rising tone, otherwise its vowel elides (see paradigms in W 2002: 26-31, 78ff. V 2004 49 ff.).	no	Present depending on the conjugation classes (V 2004: 49-59)
-7	LEX/ADV	Incoporate	Can be core lexical meaning (59). incorporates can include nouns or adjectives	X		yes	Always bears rising (or rising and falling) tone (see paradigms in W 2002: 26-31, 78ff. V 2004 49 ff.).	no	Present depending on the conjugation classes (V 2004: 49-59)
-6	P/N/G	Subject/ object	ba/bo (1sg) a/o/bu (3sg.M) Ø/i/u/bu (3sg.F/N) dəŋ (1pl) kəŋ (2pl) aŋ/oŋ/bu (3pl) (V 2004: 48)	L	See description above.	pos.	Can bear rising or falling tone (the latter only when P-7 incorporated elements occur) (see paradigms in W 2002: 26-31, 78ff. V 2004 49 ff.).	no	Present depending on the conjugation classes (V 2004: 49-59)
-5	LEX/ADV	Thematic, adverbial	labialization+k (move away, down, off, comitative, change of state) k (proximal-deictic, dynamic introverted motion) h (long/straight object classifier) t (superficial contact with surface/classifier for mental states) q (lying face down/inessive/causative) n (circularity, repeated action/head/roundness) ŋ (using one's own eyes) (V 2004: 62-64)	X		(pos.)	Coda of P-6, can occur as first segments in words, e.g. t-á-b-à-daq (TH-PRES-3sg.N-3sg-fall) 'it falls.' (Vajda 2004: 55)	no	Present depending on the conjugation classes and valency (V 2004: 49-59)
-4	ADV/T/P/N/G	Object, present, past, resultative	a (PRES/Resultative) s/i (PRES) o (PAST) a (3sg.M.PRES) o (3sg.M) i (3sg.F) aŋ (3pl) oŋ (3pl.PAST) (V 2004: 48, 68) (See NV 2015: 34 for choosing tense categories instead of aspectual categories)	L	See description above.	pos.	Can bear rising or falling tone (see paradigms in W 2002: 26-31, 78ff. V 2004 49 ff.).	no	Present depending on the conjugation classes (V 2004: 49-59)

KET (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-3	ADV/V/G	Subject/ object, applicative, pluractionality	b (3sg.N/applicative/Pluractionality) (V 2004: 48, 68,72)	L	See description above.	(pos.)	Either Onset or Coda (see paradigms in W 2002: 26-31, 78ff. V 2004 49 ff.), can occur as the last segment of the word such as in di-p 'I eat it.' (Vajda 2004: 11)	no	Present depending on the conjugation classes (V 2004: 49-59)
-2	T/M	Past, imperative	il/in (IMP/PAST) (V 2004: 48)	L	See description above.	pos.	Can bear rising or falling tone (see paradigms in W 2002: 26-31, 78ff. V 2004 49 ff.).	no	Present depending on the conjugation classes (V 2004: 49-59)
-1	P/N	Subject, object	di (1) ku (2) a (3) daŋ (1pl) kaŋ (2pl) aŋ (3pl) (V 2004: 48)	L	See description above.	pos.	Can bear falling tone (see paradigms in W 2002: 26-31, 78ff. V 2004 49 ff.).	no	Present depending on the conjugation classes (V 2004: 49-59)
0	LEX/ADV/V	Root, Trasitivity, iterative, semelfactive	Some Roots have suppletion for number. Some roots are used for valency, like t (transitive) ges (semelfactive) (72) qan/saŋ (in combination with k in P-5) (inchoative) (73)	X		pos.	Can bear falling tone (see paradigms in W 2002: 26-31, 78ff. V 2004 49 ff.).	(yes)	"Every verb fills P0 with what is called a "base", though in rare cases the entire base elides phonologically." Copare di-p (1.subject-3.object) 'I eat it' (59)
1	N	Subject	(i)n (pl) (V 2004: 48)	L	See description above.	pos.	Can bear falling tone (see paradigms in W 2002: 26-31, 78ff. V 2004 49 ff.).	no	Present depending on the conjugation classes (V 2004: 49-59)

KIOWA (LCI sample); Watkins (1980, 1984)

Prominence: high or falling tone.
"Syllables with high and falling tone are more stressed than those with low tone, a long syllable is more stressed than a short one, length combined with high or falling tone makes stress particulary marked." (49)

Obligatoriness: "The obligatory elements are the stem, a pronominal prefix and an inflectional or modal suffix" (188)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-3	P/N	Subject/object	Intransitives subjects: à (1sg) è (1pl/3 inverse) èm (2sg) mà (2du) mà (2pl) bà (2pl) Ø (3sg) ę̀ (3du) á (3pl) Transitive subjects with singular object: gyà (1sg) é` (1pl) à (2sg) má`(2du) bá` (2pl) Ø (3sg) ę́` (3du) á`(3pl) é` (3inv) (165) etc. (see pp. 148-149 for full paradigms).	G	Form determined by animacy and number of the acting scenario, also classes based on transitivity (172)	pos.	Patient prefixes have all high tone (149) whereas intransitive and agent prefixes high and low tones. (148)	(yes)	"Obligatory first element in the verb complex" (141) but can be zero in third person intransitive (165)
-2	ADV	Incorporation	Different adverbs (very, often, again, etc.) (278)	X		pos.	Can be low, falling and high tone (see pp. 278)	no	
-1	LEX	Incorporation	verb stems (277) different nouns (279)	X		pos.	Can be low, falling and high tone (see pp. 276-284)	no	
0.1	LEX/A/M	Root, imperfective, perfective, imperative	See vowel alternation for perfective (206, 209, 210) and Imperative (215). Default root is imperfective root.	L	Root alternation only to a group of verbs (206 for perfective) and (215 for imperfective)	yes	All the imperfective, perfective and imperative stems seem to be falling or high (206-218).	yes	
0.2	LEX/V/A	Root, perfective, imperfective, stative	n/l or root-final consonant (imperfective) (200) l (Stative) (194) root final consonant t/m/n/y/p (perfective) (209)	L	The appearance whether -l, -n or stem consonant is used is not predictable. "The majority of verbs has an imperfective stem form that serves as the base for imperfective inflectional suffixes" (200)	(yes)	Always consonantal, can be at the end of the word and part of prominent (high or falling tone) syllable.	yes	Since stems either end in a consonant or, if not, imperfective/perfective allomorphs occur, this position is always filled. See p. 200 for more clarification.

KIOWA (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
1.1	V	Transitive, stative, causative	y (transitive, some verbs) m (transitive, some verbs) gé/gyá (productive derived detransitive) (192) mé/bé (stative) (193) dɔ´ (stative) ɔ´m (causative) (194), bé/dé/gé thematic stems (198)	X		yes	All the valency allomorphs have a high tone.	no	
1.2	A	Perfective	Ø/ɔ́ (tr)/iá(y) (itr)/é/ (perfective 206).	L	Perfective allomorphs are not predicable: "Many roots ending in a nasal, glide, or vowel have a zero perfective suffix." (206) "A few verbs belonging to each of the preceding classes, all of them transitive, have the perfective suffix é" (208) "A number of roots with final vowels have perfective forms with a final consonant. The particular consonant (m, n, y, p) is not predictable and appears to be a relic of Proto-Kiowa-Tanoan" (208)	yes	All the perfective allomorphs have a high tone.	no	
1.3	A	Imperfective	Cɔ`/gù (transitive imperfective) (m)ià (intransitve imperfective) p (thematic stems) (211) î (imperfective imperative/future) (202) mà (intransitive imperfective) (212)	G	Imperfective allomorphs dependent on transitivity (211) but also whether the root has a thematic extension (usually verbalizer) (214). A larger number of verb stems are thematic verbs derived usually rendering causative events (198) .	pos.	Imperfective imperative suffix -î has a falling tone (202), all other imperfective morphs have a low tone (211ff.).	no	
2	PO	Negative	(C)ɔ^ (Negative) (225) gû: (transitive active negative) yɔ^ (intransitive active negative) (226) gɔ^ (intransitive stative negative) (227)	G	Negative uses perfective root (202). Negative allomorphs according to transitivity (225-227)	pos.	Can have high tone, see allomorphs.	no	
3	T	Future	t(')ɔ´/ t(')ɔ` (Future) (218)	G	tɔ´is used with intransitive verbs and t'ɔ´with transitive verbs. Tone allomorphy phonologically conditioned (218).	pos.	Can have high tone, see allomorphs.	no	
4	M	Hearsay (for position after future see example 189c. on p. 228)	hêl/hél (hearsay perfective) (223) dê/ê (Imperfective hearsay) (224)	G	Imperfective hearsay is -dê when affixed to stative stems or the future affix and -ê otherwise (225). The allomorphy -hêl/-hèl is phonologically conditioned (223)	pos.	Can have high tone, see allomorphs.	no	

KOASATI (LCI sample); Kimball (1991);

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-9	LEX	Incorporated noun	a:ti/a:t/at/a (someone) na:si/na:s/nas (something) (150)	X		no		no	
-8	ADV	Directional	oht/oh/o (to go and do something) i:t/it/i (to come and do something) (147)	X		no		no	
-7	V	Instrumental	s/st (instrumental) mat/mas/mast (instrumental indicating distance form the speaker and physical qualities of an object acted upon) (140)	X		no		no	
-6	ADV/V	Distributive	ho/oh (distributive), also derives passives; 136) hoho/ohoh (iterative) (139)	X		no		no	
-5	P/N	Indirect Object/ intransitive subject (when the subject is a possessed verb), direct objects (am-set) (121)	am (1sg) cim (2sg) im (3) kom (1pl) hacim (2pl) ilim (reflexive) ittim (reciprocal) (131)	L	"The am-set is used idiosyncratically with a number of transitive verbs to cross-reference a direct object, instead of the usual position 4 prefixes." (132)	pos.	These suffixes are accented in the verb "to give" (since there is no root), (104)	no	
-4	P/N	Direct object/ intransitive subject (ca-set)	Direct objects: (ca-set): ca (1sg) ci (2sg) Ø (3) ko (1pl) haci (2pl) ili (reflexive) itti (reciprocal) (127-128)	L	"Formerly, the choice of the ca-set of pronouns or am-set to inflect a stative verb was semantically governed, with ca-stative forms being used for states considered more or les permanent, while am-statives were used for states considered more temporary." (253) "At present, the majority of stative verbs can be inflected with only one of the two sets of prefixes, and the choice of prefix set must be marked in the lexicon with the verb, as the choice is not semantically predictable" (253)	no		no	

KOASATI (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-3	ADV	Specific locative	itta (action on the ground) (116) o:/o:w (action in water (118) pa (action on a raised, arificial,or non-ground surface) (119) on (on a vertical plane) (121) itta (in the middle of) (123) ibi (on the human face) (123) ico: (on the human mouth) (124) no (action on the human throat) (125) nok (action in the human throat) (126) ac (motion outwards) cok (motion inwards) ako (motion downwards) (127)	X		no		no	
-2	ADV	General locative	a (general locative) (115)	X		no		no	
-1	P/N/PO	Subject	l (1sg) ak (1sg.NEG) is (2sg) cik (2sg.NEG) Ø (3) ik (3.NEG) il (1pl) kil (1pl.NEG) has (2pl) hacik (2pl.NEG) (114) (can also appear as infixes 61 f.)	L	The am-set is used to cross-reference the subject of certain stative verbs (132)	pos.	Accented in some verbs, like ís-m (you gather) (58)	no	While subject morphs are obligatory, the filling of a speific slot is not, since they occur, depending on conjugation class, either as prefixes (P-1) or suffixes (P1) (see p. 58-89)
0	LEX/ADV/A/M	Root with internal changes (250ff.): indicative, aorist, interrogative, h-grade, (sequential) n-grade (completely/continuing), intensive	lengthening and accenting (low or high) (indicative) (295). Accenting of the final vowel of the verb base (Aorist) (299) Infixation of glottal stop (Interrogative) (301) h (actions that occur in sequence; infix) (302) nasalization of the vowel of the penultimate syllable of the verb root and accenting (utterly/completely with verbs of state or description) (keep on, continue with verbs of action) (307) lengthening and accenting with high rising-falling accent the rightmost accented syllable of the verb + indicative base (intensive) (208)	L	Changes are not always predictable e.g. indicative: has high accent on some verbs, in others it is low (295).	pos.	"The high and low accents in verbs are induced by the Indicative form of the verb. Most verbs take the high accent in the Indicative. Only a few take the low, and these may be archaic forms. The first form induces only a high accent, while the h-grade and n-grade seem to follow the Indicative. The intesive induces the high rising falling accent." (27)	(yes)	Root does not exist in the verb "to give" and "to say" (95)

KOASATI (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explana-tion Prom.	Obl.	Explanation Obl.
1	PO/P/N	Subject/ polarity	(V)l(i) (1sg) o/ọ (negative) (58) lahọ (af-firmative.stative) olahọ (negative.sta-tive) (59) i (1sg) (61) c(i) (2sg) Ø (3) (h)il(i) (1pl) hác(i) (2pl) tákkọ (1sg.NEG) cíkkọ (2sg.NEG) kọ (3.NEG) kílkọ (1pl.NEG) hacíkkọ (2pl.NEG) (63, 64) lí (3) (69) (h)ís(k) (2sg) (79,81) ás(k) (2pl) ílk (1pl) (82) tíska (2sg) tílka (1pl) táska (2pl) (84)	L	The allomorpy of these suffixes depends on their s(ub)class (see p. 58-89)	pos.	Usually ac-cented, ex-cept for first person singular al-lomorphs (See para-digm clas-ses 58-89).	no	While subject morphs are obligatory, the filling of a speific slot is not, since they occur, depending on conjugation class, either as prefixes (P-1) or suf-fixes (P1) (see p. 58-89)
2	ADV	Adverb	á:ho(:)s(i) (very) báhno (must; be obliged) fíhn(a)/fį/fin (too much) hónk(a) (really) má:li (in the same way) mbí:k(a) (a pleasant deal) V´:mo (in-tensely) ná:n(a) (all the time) palámmi (extremely) (158)	X		yes	All suffixes are ac-cented (158).	no	
3	ADV	Diminui-tive, inten-sive	o:si (a little; completely) (159)	X		no		no	
4	A	Habitual	v´hco (habitual) ´vhco:li (customary) v´hci (progressive) (159)	U	No conditioning found (175 - 180)	no	This suffix is never ac-cented but induce an accent in the vowel to which they are at-tached (175).	no	
5	M	Intention	áhi (intention) á (immediate intent) (159)	U	No conditioning found (180 -183). The following uses of the suffixes are all related: "this suffix with the same-subject switch-reference marker -k is used in constructions as an alternative to the verbal noun. áhi + iterrogative questions: polite questions; áhi + má:mi 'dubitative': 'to be ready to'. áhi + tense suf-fixes: actions that had the potential to be fulfilled, but were unrealized. The suffix 'á' usually occurs alone. (Kimball 1980: 161)	yes	All suffixes are ac-cented (159).	no	

KOASATI (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
6	ADV	modal adverbs	bá:no (regularly; occasionally) bí:no (reluctantly, shyly) /halpí:s(a) (to be able to) yáhl(i) (to be obliged to; really) (159)	X	No conditioning found except phonological conditioning (Kimball 1980: 161ff.)	yes	All suffixes are accented (159).	no	
7	ADV	Realis/Irrealis, Future	laho (irrealis:future) laha (irrealis:emphatic future) toho (realis; preterite) toha (Realis; perfect) (159-160)	U	No conditioning found (190-194 ff.)	no		no	
8	ADV	Deduction	o:li (facts about action are deduced by speaker) (160)	X		no		no	
9	ADV	Modal adverbs	á:p(i) (almost) áhp(i) (merely; just) má:l(i) (must/would) :sahawa (must be/ might be) yá:li (occasionally; superficially) (160)	X		pos.	only :sahawa is unaccented (160)	no	
10	ADV	Dubitative adverbs	má(:m(i)) (perhaps; maybe) má:s (dubitative of observation) (160)	X		yes	All suffixes are accented (160)	no	
11	ADV	Hearsay	mp(a) (reported speech, first or second hand) (160)	X		no		no	
12	ADV	Auditory	ha(wa) (Auditory evidence of action ocurring) (160)	X		no		no	

KOASATI (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
13	ADV/T/M	Tense, conjunction, negative imperative	:s(a) (Past I) t(i) (Past II) t(o) (Past III) k(i) (Past IV) :fó:k(a) (when; while) V´n(na) (negative imperative) (161)	U	No conditioning found except phonological (207-212)	no		no	
14	M	Consequence, subjunctive, but	:p (if, when' Subjunctive I) :k (if, when' Subjunctive II) sk(a) (because) y (but, contrary to expectation) tik(a) (but) (161)	U	No conditioning found except phonological (213-220)	no		no	
15	ADV	Phrase terminal marker, switch reference marker, question etc.	Ṿ (vowel nasalization) (phrase terminal marker) deletion (phrase terminal marker deletes word-final vowels) ś (phrase terminal marker; obsolete marker of 'male speech') n (switch reference marker, different subject) k (switch reference marker, same subject) on (switch reference marker, different subject; focus) ok (switch reference marker, same subject; focus) p (new topic) t (verb connector, coordinating) h (verb connector, subordinating) oʔli (question suffix) oʔló (tag question suffix) V´ʔwí (desiderative tag question suffix) V´ʔhá (either/or question suffix) háʔwá (rhetorical question suffix) Ṿh (delayed imperative suffix) (161-162)	X		no		no	Deletion (221) and nasalization (222) (terminal markers) do not add segmental material to this slot.
16	ADV	but it is the case that, it is similar action to (clitic)	katík (but it is the case that) máhco (it is a similar action to) (162)	X		yes	All clitics have accent (235)	no	

KOREAN (GCI sample); Sohn (1999)

Prominence: stress
"Non-phonemic stress normally occurs on the first syllable of an intonational phrase, especially when that syllable ends in a consonant; phonemic long vowels, which are allowed only in phrase-initial syllables, receive stress. If the phrase-intiial syllable ends in a simple vowel and the second syllable has an onset, the second syllable tends to receive stress. (...) on the other hand, a speaker may put stress on any word which he thinks is relatively important or needs to be emphasized or focused (197).

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-1	ADV	Derivation	Several adverbial affixes, e.g ch(y)e (recklessly) cis (randomly) es (crookedly) nay (outwardly)nuc (late) pi (twisted) pis (aslant) sel (insufficiently) tes (additionally) yes (stealthily) (224)	X		pos.	Can be in phonological phrase initial position, and therefore stressed, such as in Pak-sensayng-nim i tólaka-sy-ess-ta-ko (Park-teacher-honorific nominative pass away-go-honorific-past-declarative-question) '(Are you saying that Professor Park passed away?) (200)	no	
0	LEX	Root		X		pos.	Can be in phonological phrase initial position, and therefore stressed, such as in swúm-ca hide-propositive.plain (Let's hide) (202)	yes	
1	ADV/V	Derivation	chi (transfer) i/hi/li/ki (thematic/valence affixes; causative, passive; dependent on the verb) (225) cilu (intensifier) coli (gently) kkali (intensifier) m (intensifier) ttul (intensifier) ttuli (intensifier) ci (get characterized by, denominal) i/hi/chwu/iwu (deadjectival causative) ci (get characterized by, deadverbial) i (be doing, deadverbial) keli (keep doing, deadverbial) tay (do, repeatedly, deadverbial) (225-226)	X		no	Never at the beginning of the phonological phrase, since verb roots must precede (233).	no	

KOREAN (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
2	HO	Subject honorific	si/sy/usi/usy (Subject honorific) (233)	U	No conditioning except phonological (233)	no	Never at the beginning of the phonological phrase, since verb roots must precede (233).	no	
3	A/T	Past, present, perfect, past-past	ass/ess (past/present perfect) assess/essess (past past/past)	U	No conditioning except phonological (233, 362–366)	no	Never at the beginning of the phonological phrase, since verb roots must precede (233).	no	
4	M	Intention, conjecture, prospective	keyss (intention or conjecture) ul(i) l(i) (prospective) (	U	No conditioning except phonological (233)	no	Never at the beginning of the phonological phrase, since verb roots must precede (233).	no	
5	M/HO	Addressee honorific, indicative, retrospective, requestive, declarative, interrogative, propositive, imperative (clitic)	(nu)nta (Declarative.Indicative.Plain) tela (Declarative.Retrospective.Plain) nunya/ni (interrogative.Indicative.Plain) ten(ya)/ti (Interrogative.Retrospective.Plain) Ø/e (Declarative.Intimate.Indicative/Interrogative) ney (Declarative.Indicative.Familiar) tey (Declarative.Retrospective.Familiar) (nu)nka/na (Interrogative.Indicative.Familiar) tenka (Interrogative.Retrospective.Familiar) so/(u)o (Declarative/Interrogative.Indicative.Blunt) (e/a)yo (Declarative.Indicative/Interrogative.Polite) teyyo (Declarative.Retrospective.Polite) tenkayo (Interrogative.Retrospective.Polite) (su)pnita (Declarative.Indicative.Deferential) (su)ptita (Declarative.Retrospective.Deferential) (su)pnikka (Interrogative.Indicative.Deferential) (su)ptikka (Interrogative.Retrospective.Deferential) ca (Propositive.Plain) (e)la (Imperative.Plain) Ø/e (Propositive/Imperative.Intimate) sey (Propositive.Familiar) key (Imperative.Familiar) (u)psita (Propositive.Blunt) (u)o (Imperative.Blunt) (e)yo (Propositive/Imperative.Polite) sipsita (Propositive.Deferential) sipsio (Imperative.Deferential) (236-238)	G	Morphemes depend on the following cross-cutting categories: Declarative, Interrogative, indicative, Retrospective, Propositive, Imperative, Plain, Intimate, Familiar, Blunt, Deferential (236-238)	no	Never at the beginning of the phonological phrase, since verbs or adjectives have to precede (236).	(yes)	These morphs are obligatory in the clause. Only in Intimate Honorific, when the root ends in a vowel, there is no overt marker (236-237).

LUMUN (LCI sample); Smits (2017)

Prominence: higher or falling tone.
Although Smits (2010) lists four tonemes (81), the falling and rising tones are phonetically combinations of two tones. In many contexts (83). Rising tone on a single mora occurs only in prepausal position (87). Rising tones are often pronounced as low tones (90). Furthermore, high and falling tone are acoustically higher than rising tone (95). Also, "high tone (…) mimics the stress placement in (…) Arabic (…)" (123). Given this information, it has been decided to include high and falling tone as prominent, and rising and low as non-prominent properties.

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-9	M	Conjunctive, subjunctive, restrictor, focus	á (conjunctive) â (subjunctive) í (restrictor) akk/a+CC (followed by double concord) (331)	U	No conditioning found (331)	pos.	These morphemes have all the tone varieties, only the focus prefix has no high tone (331)	no	
-8	P/N	Subject	Bound personal subject pronouns/free pronouns ń/ɔʊ́n (1) ŋ́/ɔ́ŋ/ɔʊ́ŋ (2sg) kw- + tone pattern/ɔɔ̂k (3sg) ĭt/ɔrĭt(1+2) ɪ́n/ɔnɪ́ń (1A) ʊ́n/ɔrʊ́n/ɔrɔ́n (1+2A) nń/ń/ɔ́n/ɔnɔ́n (2A) n̂- + tone pattern/ɔkîn (3A) (205)	U	"Singular subjects, when immediately preceding a verb or predicative adjective, are far more commonly expressed by a bound pronominal form" (200). "Use of the free pronoun(...), is nevertheless possible." The bound pronoun is commonly used for 1, 2 and 3 subjects, while the free pronoun is commonly used for 12, 1A, 12A, 2A and 3A subjects (206), but there is no hard conditioning. **The bound forms, however, can coalesce with the concord marker. They do not appear inimperatives and hortatives though (202)**	pos.	Only 3A 'they' from the bound pronouns has a high tone (207). Free pronouns have more diversity (198)	(yes)	The subject can also be a noun phrase instead of a bound pronoun, therefore, there is a zero morpheme for third person (206).
-7	P/G	Subject	p/t/t̪/c/k/kw/m/n/ɲ/ŋ/l/w/Ø (different prefixes dependent on noun class) (189)	G	The existence of concord is conditioned by T/A/M: present in incompletive, past, completive, not present in imperative, dependent incompletive, dependent completive (338)	(pos.)	Consonantal and onset, can be at the beginning of the word and as such take tones of the nucleus of the syallable they belong to.	no	Only in finite TMA constructions
-6	ADV/A/T/M	Possibility	anta (can), inflecting for T/M/A (411)	G	Allomorphs for TAMs (405)	pos.	Can have two tones (low, high) (411)	no	

LUMUN (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-5	ADV/A/T/M	As always'	arəka (as always) (inflecting for T/M/A) (405)	X	Allomorphs for TAMs (405)	pos.	Can have all the tones (falling, low, high) (405)	no	
-4	A/T/M/PO	Negation	ɔkə́rənnɔ (let, abstain), inflecting for T/M/A (423-424)	G	Allomorphs for TAMs (424)	pos.	Can have both high and low (see alternatio-non p. 424)	no	
-3	ADV/A/T/M	Again'	app/ápp (again) (inflecting for T/M/A) (418)	X	allomorphs for TAMs (418)	pos.	Can have both high and low (see alternatio-non p. 421)	no	
-2	ADV/A	Itive, veni-tive	ɔ́t̺ (itive dependent incompletive) at̺ (ventive dependent incomplet-ive) át̺ (itive incompletive) at (ventive completive) âtt (itive completive) âtt (ventive complet-ive) t (itive imperative). No ventive imperative. (439) at̺ com-binations of auxiliaries (p. 457)	X	itive and ventive allo-morphs for different TAMs (see 439)	pos.	Can have low, rising and high tone (see forms on p. 438)	no	
-1	M	Irrealis	ɔ̂/î/û/ɔ̂,/ɔ́/â (irrealis dependent) á/â (+ concord) (irrealis independent) (431) phonologically conditioned allophones (431)	G	Irrealis allomorphs for dependent or independ-ent. Some verbal con-structions require a de-pendent form, for exam-ple negation (432)	yes	Can have high falling tone or high tone (432)	no	
0.1	LEX/A/T/M	Imperative, incomplet-ive, dep. in-completive, completive, past, dep. perfective (see 338), ir-realis (initial stem vowel) (431)	ɔ/a/ɛ/ɪ/ʊ (338) Stems that are not ɔ-initial do not change their initial vowel, and stems with another tone pattern than L.L* retain their stem tones (344)	L	Only stems with ɔ alter-nate for categories, be-tween ɔ and a, which means that a form start-ing with a- could be ei-ther an ɔ- or a-stem, and therefore the category is lexically conditioned (see paradigm on Table 50, 344)	pos.	All verbs have a low tone on the first mora (337). A stem initial vowel o- changes into a- and a high tone appears on the initial vowel of all-low stems, Some verbs with an all-low tone pattern have a falling tone on the initial vowel (338). Unless in careful speech, this falling tone can also be realized as high. Which verbs belong to this group is lexically determined. Also see incompletives carry a high tone (344)	yes	A stem al-ways has a stem vowel (333)
0.2	LEX	Root	Lexical syllable if the verb is tri-moraic; verbs are mono- bi- or tri-moraic (333)	X		pos.	Roots can have all tones, usually the second mora is high (334-335)	no	
0.3	LEX/ADV	Root, Plu-ractionality	Gemination of t, k, nasal or rho-tic/ll/Vtt/ʊkk(w)/ccɛ/partial redu-plication/homorganic obstru-ent/homorganic nasal/ɲc/ɛnt /vowel lengthening/ɛ(P2) (Plu-ractionality) (462)	X		pos.	Roots can have all tones, and the morphono-logical alternations in this position are part of the root (462-465).	no	

LUMUN (cont.)

Pos.	Catego-ries	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explana-tion Obl.
1	V	Benefactive	ɪn(tɛt)/ɛn(tɛt)/an(tɛt) (benefactive) (337) (the tɛt-part belongs to the following po-sition)	X		pos.	Usually low, but high tone in the completive (369)	no	
2	V/A/T/M	Imperative, in-completive, dep. incompletive, past, dep. perfec-tive, completive (338).	Different alterna-tions of last vowel (ɔ, ɛ or a) (333) and tone (338), e.g. i/u (imperative) (340). The final vowel can be also the final vowel of applica-tives (see 369)	L	See Table 47 for stem final imperative formation. The list does con-tain phonological information but the patterns are not always pre-dictable by the tone pattern and vowel. (340)	pos.	Different tones (high, low, fall-ing). see e.g. Table 621 on p. 367 (default verb vs. com-pletive form)	yes	There is al-ways a sec-ond syllable of the verb that con-tains the fi-nal vowel (or the vowel be-fore t) (333)
3	V/A/T/M	Completive, in-completive, de-pendent perfec-tive, imperative, benefactive, past	t (completive/in-completive/depend-ent perfective) t̪ɛ́ (imperative) t̪ɛ̂ (completive) t̪ɛ (past) (338) (can be analyzed as part of the stem; (t̪ɛ)t (ben-efactive/completive benefactive) (369)	L	"Verbs with the Locative-applicative suffix have basic TAM- mor-phology that is different from the basic TAM-morphology of vowel-final verbs. The presence (or absence) of the Locative-applicative suffix can be recognized in Imperatives, Incompletives (both non-dependent and dependent) and Completives, but not in Pasts and de-pendent Perfectives." (338). The semantics and application of the applicative -t is lexicalized in some verbs, and therefore the asso-caition of it being a t-TMA marker too (see 503). Therefore, mood and aspect is lexically conditioned in this position. Imperative seems to be t̪ɛ́ for some verbs like "send" that ends in -t (ɔt̪ɪ̪ɔ́t). It is lexically conditioned, because the suffix can mean past in other tenses; compare Table 46 in p. 340 with imperative allomorph í or Ø.	pos.	Past low tone (grammatically conditioned), Imperative high tone (lexically conditioned) (340)	no	

LUMUN (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
4	ADV/P/N	Object (214), vague reference (clitic)	ń/kín/ín (1sg) ŋ́/kóŋ/óŋ/ŋóŋ (2sg) k/kɔk/ɔk/ŋɔk (3sg) tĭt (1+2) nín (1.A) tón tɔ́n(12A) tɔ́n(2A) kɪn (3A) (216); cĭk (vague reference particle) (567)	L	"The object pronominals with singular reference (1, 2 and 3) come in four paradigms. The choice between the first three of these paradigms is conditioned by the morphology of the verbal TAM and the tone class of the verb involved" (i.e. lexically conditioned) "I call these paradigms group I, group II and group III. The forms of group I and group III in some cases interact with the preceding verb in ways that do not necessarily follow from phonological rules, and the initial k of the group II forms is epenthentic. The singular object pronouns of the first three paradigms are clitic." (215)	pos.	Can have high or low tone (See forms 216)	no	Object pronouns can be also omitted (only appear in transitive clauses)
5	ADV/M	Adverbial, question (clitics)	Adverbial particles: a (attention) (636) na (mild imperative) (637) tɪ ('you know') (638), mɛ́ (proposal for action) (639), mɛ (polite commands with urgence) (640); Question particles: ɪ (neutral polar question) (711) páɪ (is it?) (714) a (information recovery) (714) ɛ (informal information recovery) (715)	U	No conditioning found (636-640; 711-715)	pos.	Only mɛ́ and páɪ have high tone (331)	no	

LUO (LCI sample); Tucker and Creider (1994)

Prominence: stress
Although Smits (2010) lists four tonemes (81), the falling and rising tones are phonetically combinations of two tones. In many contexts (83). Rising tone on a single mora occurs only in prepausal position (87). Rising tones are often pronounced as low tones (90). Furthermore, high and falling tone are acoustically higher than rising tone (95). Also, "high tone (…) mimics the stress placement in (…) Arabic (…)" (123). Given this information, it has been decided to include high and falling tone as prominent, and rising and low as non-prominent properties.

Pos.	Catego-ries	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Expla-nation Prom.	Obl.	Explanation Obl.
-3	PO	Negation (clitic)	kík (negative imperfect/subjunctive/present/future) (474, 477) ok/pók (negative perfect aspect) (474-480)	G	Negative proclitics are dependent on mood and aspect (473 ff.)	no	Not stressed, since not part of the stem vowel.	no	
-2	T	Future, past, present (458) (clitic) "The tense particle may precede or follow the negative ok. In the latter they merge with the Personal Prefix" (479)	á-`ye (general past) (460) nê (Recent Past) (460) néné (Remote Past) (460) nyô-ro (Past of Yesterday) (461) nyóca (Past of the day before yesterday) (462) yá`nde/yandê (Recent Past, indicates the Past of a few days ago) (463); n(í) + Subjunctive stem (Future) (464) Ø (Present tense) (implied; there is no listing for present tense)	U	No conditioning found (459-464)	no	Not stressed, since not part of the stem vowel.	(yes)	Only zero marked form is present (459)
-1	A/M/P/N	Subject, imperfective/completive/perfective, indicative/imperative/subjunctive	Imperfective subject markers: á (1sg) í (2sg) ó (3sg) wá (1pl) ú (2pl) gí (3pl) (350-351) (the prefixes tone are in harmony with the one of the stem vowel (350); Perfective/completive subject markers: a (1sg) i (2sg) o (3sg) wá (1pl) ú (2pl) gí (3pl) (354-355) (The tone of the prefixes is low, but is raised if a self-standing pronoun precedes) (354) o (subjunctive) (367) Imperative reading of subjunctive does not have a prefix (358)	G	Different subject prefixes for aspect sets ("These Aspects are distinguished by Tone alone when there is a Pronominal Subject; when the Subject is a Noun, however, the Verb in the Perfect Aspect has a Prefix o-/o-" (350).	no	Not stressed, since not part of the stem vowel.	(yes)	Subject noun phrase usually elides the pronominal prefix of third person (see examples on p. 351-352). The prefixes are usually not found in subjunctive, but the suffix o- might apear occasionaly instead of it "It is possible that the presence of Prefix o-/o- indicates permission or persuasion, while its absence indicates command or complusion or obligation." (Footnote, 366). Thus, only imperative contexts (compulsion, obligation) account for zero prefixes, thus, prefixes can be seen as obligatory.

LUO (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
0	LEX/A/M	Root, Imperfective/Incompletive, Perfect/Complete Indicative Subjunctive-Imperative.	Tone alternation for: imperfective/incompletive and perfect/complete (350) indicative, subjunctive-imperative (346), contracted stems for subjunctive in some stems (379).	L	Conjugation classes are posited without semantic or syntactic correspondence: "Full" CVCV or VCV (345) Low-Tone-Contracted Stems (379) Low-Tone Short Verbs (413) Stems with Suffixes (425)	pos.	"Stress falls primarily on the Stem vowel of a word and virtually never on a Prefix or Suffix" (18). "An exception to this general rule may be found in the Plural Imperative of Verbs, where Stress falls on the suffix, e.g. ru'dúrú 'Stir' 'rudo 'Stir'!)" (Footnote 18)	yes	
1	LEX/M	Root, Imperative/Subjunctive, (stem-final element)	í (after s and h stem consonant) or stem vowel (Imperative/Subjunctive) (358). Contraction of yo in contracted verbs (379)	G	Imperative/Subjunctive elides when there is an object following, tone shifts to stem if elided (360). Contraction occurs with verbs ending in yo, but the contraction is optional, not lexically conditioned (379)	pos.	Stress can fall on this part in imperative forms (18, 358)	no	
1	V	Applicative, reciprocal	n(i) (benefactive) (347, 352) r (reciprocal) (431)	X		no		no	
2	P	Reciprocal object, (in)direct object	Object suffixes: á (1sg) í (2sg) é/Vgo (3sg) wá (1pl) ų́ (2pl) gí (3pl) (347); Reciprocal suffixes: â (1sg) î (2sg) ê (3sg, 1/2/3pl) (431)	G	Object is affected by tone in subjunctive (372)	no		no	
3	V	Instrumental, (clitic; see positioning with object and úrú: ted-gó-do-urú (cook-3sg.OBJ-INSTR-IMP.pl) or ted-úru-gó-dó (cook-IMP.pl-3sg.OBJ-INSTR))	d-ó 'with it' (353)	X		no		no	
4	M/N	Imperative plural	úrú (imperative plural) (363). -úrú can occur also before P2, but there is no hard conditioning, as it seems (363-366).	G	Plurality conditioned by mood or vice versa (-í in P 1 represents singular imperative) (363).	yes	"An exception to the general rule may be found in the Plural Imperative of Verbs, where Stress falls on the suffix, 'ru'dí 'Stir'! ru'd-úrú 'Stir-PL.IMP' (Footnote on p. 18).	no	

MAPUDUNGUN (GCI sample); Smeets (2008)

Prominence: stress
"The tendency is to have stress on the vowel before the last consonant of the word. Trisyllabic words tend to have the second vowel stressed." Longer words may have several stressed syllables. In general, every second and every last vowel is stressed. The second vowel usually has primary stress" (49

Pos.	Categories	Functions	Morphs/Features	Cond.	Explana-tion Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
0	LEX	Root, Redu-plication		X		pos.	Usually stressed when bi-syllabic; Re-duplicated elements are equally promi-nent e.g. kúykúypángi (50)	yes	
1	ADV	Verbalizers	tu (playful, non-serious) (304) nge/ye (last-ing situation in which the event takes place repeatedly and with intensitive) (305)	X		no	Markers occur after reduplicated sylla-bles, which are always stressed. (304ff.)	no	
2	ADV/V	Experience, applicative	ñma (experience) ye (oblique object) (301)	X		pos.	ñma can be in ultimate closed syllable as in poy-má-y boil-experience-3sg.in-dicative 'he got a boil'. (301); ye can be in second position of a multisyllabic word as in dhuam-yé-nie-fi-n (want-oblique.object-progressive.persistent-3.patient-1sg 'I need him.' (303)	no	
3	V	Causative	(ü)m causative (ü)l causative (299)	X		pos.	(ü)m can be in second position of a multisyllabic word as in ñam-üm(...) 'to lose' (299) el can be in second posi-tion of a multisyllabic word as in af-el-(...) 'to end' (tr.) 'to be bored with'. (300)	no	
4	ADV/V	Transitiv-izer, causa-tive	tu (transitivizer) ka (factitive) (297)	X		pos.	–tu can be in second position of a mul-tisyllabic word as in ñam-üm(...) 'to lose' (299) el can be in seod position of a multisyllabic word as in af-el-(...) 'to end' (tr.) 'to be bored with'. (300)	no	

MAPUDUNGUN (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
5.1	A	Progressive persistent, perfect persistent	nie (progressive persistent) (293)	G	"A telic verb which contains the suffix -nie- or the suffix -künu is result oriented. An atelic verb which contains one of thee suffixes is event-oriented." (294).	pos.	Can be in ultimate closed syllable as in küdhaw-nie-n ñi mapu work-Progressive.persistent-1sg.indicative poss.1sg land 'I cultivate my land' (i.e. it does not lie fallow. The subject is not necessarily working the land at the moment of speaking.' (295)	no	
5.2	A	Perfect persistent	künú (perfect persistent) (293)	G	"A telic verb which contains the suffix -nie- or the suffix -künu is result oriented. An atelic verb which contains one of thee suffixes is event-oriented." (294). "The perfective form of posture verbs obligatorily contains the suffix künu" (296)	yes	künú always bears stress (50)	no	
6	V	Reflexive/ reciprocal	(u)w (reflexive) (290)	X		pos.	Can be in second position of a multisyllabic word as in traw-úw-üy-ng-ün gather-reciprocal-indicative-3pl.indicative 'They gathered together' (209)	no	
7	ADV	Circular Movement, intensive	(k)iaw (circular movement) (kü)tie (intensive) (288)	X		pos.	Can be in penultimate syllable followed by an open syllable as in kishu-yaw-chi alone-circular-subjective.noun 'go about alone' (376-377). No evidence found for (kü)tie being in a stressed position.	no	
8	ADV	More involved object	(ü)l (more involved object) (287)	X		pos.	Can be in second position of a multisyllabic word as in maychü-l-fi-ñ wave-more.involved.object-3>3-1sg.indicative 'I signaled to him' (288)	no	
9.1	A	Progressive	meke (progressive) (286)	U	No conditioning found (286)	pos.	Can be in ultimate closed syllable as in añütu-meki-y be.drowsy-progressive-3sg.indicative 'he is drowsy.' (286)	no	

MAPUDUNGUN (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
9.2	A	Stative, progressive	(kü)le (stative) (280)	G	Telic verbs with statives are result-oriented; with atelic verbs, the stative suffix denotes an ongoing event (281) or, among other functions, a subject still being capable of performing the action denoted by the verb (280)	yes	(kü)lé always bears stress (50)	no	
10	V	Applicative	(l)el (beneficiary) (278)	X		pos.	Can be in second position of a multisyllabic word as in ye-lél-fal-ma-nge-y-mi mi maleta carry-benefactive-force-indirect object-passive-indicative-2sg.indicative 'they have to carry your suitcase for you' (279)	no	
11	V	Applicative	(ü)ñma (indirect object) (276)	X		pos.	Can be in second position (in trisyllabic words) as in i-ñmá-nge-y kiñe sanchu ñi karukatu eat-appl.indirect_object-passive-3sg. one pig poss1sg neighbour 'They ate a pig of my neighbour's (lit. my neighbour was eaten one pig)'. (276)	no	
12	ADV	Force, satisfaction	fal (force) (ü)ñmu (satisfaction) (272)	X		pos.	Can be in second position of a multisyllabic word as in ye-fal-el-nge-me-y-mi carry-force-benefactive-passive-thither-indicative-2sg 'you have to be brought bread' (275) (ü)ñmu can be in last closed syllable, as in küdhaw-tu-ñmu-n work-verb-satisfaction-1sg.indicative 'I have done my best' (275)	no	
13	ADV	Pluralizer	ye (Pluralizer)	X		pos.	Can be in second position (in trisyllabic words) as in ye-yé-pa-n bring-plurality-hither-1sg.indicative 'I bought many [things]' (272)	no	
14	V/P	Passive, agent	nge (passive) (u)w (1st person agent) mu (2nd person agent) (267)	G	(u)w: "The first person agent marker is used when the total number of participants is greater than two." (268) mu: only if the subject is first person. (268)	pos.	Can be in ultimate closed syllable as in nü-nge-y take-passive-3sg.indicative 'it was taken' (267). Mu can be in ultimate closed syllable as in lang-üm-mú-n die-causative-2.actor-1sg.indicative 'you killed me'. uw can be i ultimate position like in mütrüm-uw-lu eymün call-1.actor-subjective.verbal.noun 2pl 'when we called you' (269)	no	

MAPUDUNGUN (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
15	ADV	Play, simulation	kantu (play) faluw (simulation) (264)	X		yes	These two morphemes always bear stress (50)	no	
16	ADV	Immediate, sudden	fem (immediate) rume (sudden) (263)	X		pos.	rumé always bears stress (50), fem possible stress if it appears in second position, like ye-fém-nge-pa-y carry-immediate-passive-hither-3sg.indicative (263)	no	
17	ADV	Thither	me (thither) (261)	X		pos.	Can be in ultimate closed syllable as in küdhaw-me-n work-thither-1sg.indicative 'I went to work' (261)	no	
18	ADV	Persistence	we (persistence) (260)	X		pos.	Can be in ultimate closed syllable as in fe-m-nge-wé-n ta-tí become.like.that-causative-passive-persistence-1sg.indicative the-the 'that's how I am now.' (260)	no	
19	ADV	Interruptive	(ü)r (interruptive) yekü (interruptive) (258)	X		yes	yékü always bears stress (50)	no	
20	ADV	Hither, locative	pa (hither) pu (locative) (257)	X		pos.	Can be in ultimate closed syllable as in müle-pu-y be-locative-3sg 'live in' (258)	no	
21	ADV	Repetitive/restorative, Continuative	tu (repetitive/restorative) ka (continuative) (254)	X	No conditioning found (178)	pos.	Can be in ultimate closed syllable as in witra-tripa-tu-y get.up-leave-repetitive-3sg.indicative 'he got up and left [for home].' (255) and in: pichi-ka-y small-continuative-3sg.indicative 'he is small (as for his age).' (256)	no	
22	T	Pluperfect	(ü)wye (pluperfect) (254)	U	No conditioning found (255-256)	pos.	Can be in penultimate syllable followed by an open syllable as in aku-wyé-l-mi arrive-pluperfect-conditional-2sg.conditional 'if you had arrived' (254)	no	
23	A	Constant feature/ situation may continue into the present/ general rule	ke (constant feature/iterativity, habituality) (251)	U	No conditioning found (254)	pos.	Can be in ultimate closed syllable as in podh-küle-ké-y dirty-stative-constant.feature-3sg.indicative 'it is always dirty'. (251)	no	
24	ADV	Proximity	pe (proximity, event in the recent past) (247)	U	No conditioning found (178)	pos.	Can be in penultimate syllable followed by an open syllable as in kwida-kulliñ-pé-yu tend-cattle-proximate-1du.indicative 'We have been tending cattle' (248)	no	
25	ADV	Reportative	(ü)rke (reportative) (246)	X		pos.	Can be in ultimate closed syllable as in müle-we-rke-y be-persistence-reportative-3sg.indicative 'appeared to be' (247)	no	
26	M	Affirmative	lle (Affirmative) (245)	U	No conditioning found (245-246)	pos.	Can be in ultimate closed syllable as in chem-mew am fe-m-llé-n what-instrumental particle become.like.that-constant.feature-affirmative-1sg.indicative 'why shouldn't I have?!" (245)	no	

MAPUDUNGUN (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
27	PO	Negation	la (negative.indicative) ki (negative.imperative) nu (negative.condi-tional/subordinate/nomi-nal) (243-245)	G	Negative allomorphs are grammatically conditioned by mood and clause type (243-245).	pos.	Can be in penultimate syllable followed by an open syllable as in la-le-lá-y (die-stative-negative-3.indicative) 'he/she is not dead.' (243)	no	
28	T	Non-realized	a (non-realized situation, future, intention) (235)	U	No conditioning found that would change its meaning drastically, but nevertheless has a variety of similar meanings in different constructions (235-243)	pos.	Can be in penultimate syllable followed by an open syllable as in puw-á-yu arrive-non-realized-1du.indicative 'We will arrive' (236)	no	
29	A	Impeditive	fu (impeditive; unsuc-cesful realization) (230)	U	No conditioning found that would change its meaning drastically, but nevertheless has a variety of similar meanings in different constructions (230-234).	pos.	Can occupy the last closed syllable, e.g. with 1sg.indicative fol-lowing: müleke-fú-n (I lived) (232)	no	
30	T	Pluperfect	mu (pluperfect) (229)	G	Only combines with suffix -m (instrumental verbal noun) or -fiel (transitive verbal noun suffix). Otherwise (ü)wye (plusquamper-fect) is used (P22).	pos.	Occupies the last closed syllable in com-bination with -m (P33), however not when fiel (P33) fol-lows (230).	no	
31	P/N	Inverse/direct	fi (direct) e (inverse) (226). This position has been considered inflec-tional since it forms an integral part of Sub-ject/Object marking in P33.	G	Depends on which person is the subject. If 3rd is patient (with high animacy), then fi occurs, otherwise Ø (low animacy third per-son). If speech act participants are patient, then e is used. (226). Also, -fi and -e occur with certain subordinates in which they do not have this scenario marking function. (227).	pos.	–fi is always stressed (Zúñiga 2007: 64)	no	

MAPUDUNGUN (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
32	A	Constant feature	ye (constant feature, every time when event A occurs, event B occurs) (225)	U	No conditioning found (225-226)	pos.	Especially in combination with -m (instrumental verbal noun) (P34) it can occupy a final closed syllable, ergo stressed syllable (225)	no	
33	ADV/M/P/N	Verbal nouns, indicative, conditional, imperative, Subject/Object. This slot has been divided into three slots in Smeets (see chart on p. 181), but the interdependency of these slots and the existence of fusional morphs such as (ü)n '1sg.IND' (181) suggest a unified analysis.	Indicative: (ü)n (1s) yu (2sg) yiñ (1pl) (ü)ymi (2sg) (ü)ymu (2du) (ü)ymün (2pl) (ü)y (3sg) (ü)yngun (3du) (ü)yngün. Conditional: li (1sg) liu (1du) liyiñ (1pl) (ü)lmi (2sg) (ü)lmu (2du) (ü)lmün (2pl) (ü)le (3). Imperative: chi (1sg) nge (2sg) mu (2du) mün (2pl) pe (3) (181). Non-finite: (3pl) (ü)n (plain verbal noun) (192) el (Objective verbal noun) (200) m (Instrumental verbal noun (206) fiel (Transitive verbal noun) t (agentive verbal noun) lu/Ø (Subjective verbal noun) (ü)wma (Completive subjective verbal noun) (ü)y (Indicative except 1sg) (ü)l (Conditional) Ø (Imperative)	G	Grammatically conditioned depending on mood (P33) (181). Subject/Object status depends on the specific morpheme in P31.	pos.	Morphemes with closed syllables are stressed (the ones ending in -n or ñ, see Table 181)	yes	All suffixes are obligatory, in finite (181) and non-finite clauses (188)
33	P	3rd person agent	(m)ew (third person agent) Ø (first or second person singular agent) (178)	G	Only with speech-act participants as subjects or proximate participants acting on obviative participants (178)	yes	Always closed syllable at the end of the verb, therefore always stressed (178)	no	

MIAN (LCI sample); Fedden (2011)

Prominence: high and high-low tone.
"Syllables which are assigned H or HL from any of the tonal melodies H, HL, or LHL are more prominent than syllables which are assigned L." (83). All accented syllables have a high tone (either LHL or HL) (74)

Obligatoriness: Only few stems take prefixes (265-267)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-1	P/N/G/C	Class (intransitive subjects, transitive objects) and object prefixes	Object prefixes: na (1sg) ka (2sg) a (3sg.M) wa (3sg.F) a (3sg.N1) wa (3sg.N2/3) ya (animate plural (266); Classificatory prefixes: nem (1sg) kem (2sg) dob/do (M-class) om (F-class) dol/dl/do (1/2/3pl) tob/to (long object) gol/go (bundle-like object) gam (covering object) ob/o (residue class) tebel/tebe (long object plural) gulel/gule (bundle-like object plural) gemel/geme (covering object plural) ol/o (residue class plural) (186)	L	Seven verb stems take an obligatory pronominal prefix which indexes the object and signals person, number and gender in third person (265). Classificatory affixes show absolutive alignment (S/O) and only a "sizeable subset of the verbal vocabulary requires a classificatory prefix (267)	pos.	Prefix can bear high tone in stem accented verbs. See example 119 where tob '3sg.long.object' has a high tone (77)	no	Only few stems take prefixes (265-267)
0	LEX/A/N	Root, inanimate object number, perfective/imperfective stems.	Apophony of stem vowel (268)	L	Five perfective-only cut and break verbs indicate the number of the (inanimate) object through stem apophony e.g. wà 'cut singular object' wè 'cut plural objects' (266)	pos.	High tone occurs if the stem is accented (74)	(yes)	transfer/take' verb stem is segmentally zero (271)
1	LEX/V/A	Applicative/auxiliary -ub 'give', imperfective/perfective stem endings	ka/ê' (Imperfective) lâ' (perfective) la (perfective) Ø (perfective/imperfective, i.e. the bare stem) some irregular forms. (248) ûb' (applicative perfective) (p. 273)	L	The suffixal stem endings for imperfective/perfective containing lexical information are not predictable (248). Some verb conjugation necessitate the auxiliary imperfective bl-, others have an imperfective affix b- (p. 259)	pos.	High tone in off-stem accented verbs: verbs in which the accent is placed on the next syllable to the right of the stem, only if there are suffixes (74). For example the verb hala with the ending -lâ (Perfective) has high tone on lâ (76).	no	

MIAN (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
2	P/N/G	Recipient/object suffixes (270). The order between slot 2 and slot 3 is not mentioned, but can be induced in the chapter of object suffixing (269ff.)	ne(n) (1sg) ke(n) (2sg) a(n)/ha (3sg.M) o(n)/we (3sg.F) ye/e(n)(1/2/3 plural) (271)	L	Different allomorphs for different lexically conditioned paradigms, also some verbs cannot form objects (usually intransitives can be derived to express objects) (272-273).	pos.	Only the object allomorphs where the root is zero have high tone, like in 'give' (floating high tone) (e.g. on p. 272: monî-o om-Ø-(High tone)wen-s-e=a (money-N2 3sg.F.classif.Obj-give.PERF-3sg.F.Recip-Different.Scenario.Sequence-3sg.M.SBJ=MED) 'He gave her (a coin/bill of) money'.	no	
3	A/T/M	Near past, deontic, irrealis, realis, non-hodiernal past, habitual (auxiliary), sequential, imperfective (283) (See p. 278 example for position after the object)	n/Ø (Realis) nab (near past) b + high tone (non-hodiernal past) s (remote past) aa(m) (deontic) aamab/omab (irrealis) mab/omab (irrealis) b/l (imperfective) (283)	G	Irrealis that has two different allomorphs dependent on whether the stem ends in a consonant and is perfective or imperfective. This makes this specific allomorph grammatically conditioned. (283)	pos.	Irrealis allomorphs – the ones that are grammatically conditioned – can have high tone in off-stem accented verbs (79).	no	"Verb stems are either directly inflected for various TAM categories or they enter a periphrasis-like construction in which they are serialzed with an existential auxiliary that bears the inflectional suffixes."(283).
4	P/N/G	Subject suffixes (262)	i (1sg) eo/ebo/eb (2sg) e (3sg.M/N1) o (3sg.F/N2/N3) uo/obo/ob/bio (1exc/inc) io/ibo/ib (2/3pl) (262)	U	No conditioning found except phonological (262).	pos.	High tone when a non-hodiernal past marker precedes (305).	yes	"All finite verbs obligatorily have a pronominal suffix which indexes the subject and signals person, number and in the third person also gender of the subject" (262). Morphology is also obligatory in medial verbs (424).

MIAN (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
5	T	General past, Hesternal past	bio (general past) so (hesternal past) (296)	U	bio co-occurs with realis marker (296), but its meaning or occurrence is not determined by it. su can only occur in the perfective and the realis suffix (297) but its meaning or occurrence is not determined by it.	no	All morphemes in this slot have low tone (296).	no	
6	PO	Negation (clitic)	ba (negation) (471)	U	No conditioning found (475-476)	no	All clitics have low tone (471)	no	Clitics appear only on the final verbs (471).
7	M	Illocutionary force particles (clitic)	be (declarative) ble (exclamative) bo (Quotative, emphatic) a (Question) e (Content Question) o (Hortative 1/3 person) e (Hortative 2nd person) (471)	U	No conditioning found (472-475)	no	All clitics have low tone (471)	no	Clitics appear only on the final verbs (471).

MURRINH-PATHA (LCI sample); Walsh ("W") (1976); Nordlinger ("N") 2010; Nordlinger & Caudal ("NC") (2012); Mansfield ("M") (2017)

Prominence: stress
"All simple polysyllabic words have penultimate stress, while an internal suffix [-3–-1] causes the stress to be reassigned, but an external suffix [P1-4] has no effect on stress." (M 362) This means that only the prefixal plus stem part can have stress, and within this 'phonological word' (M 362) the penultimate syllable has stress.

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-4	LEX/A/T/PO/P/N	Auxiliary root, tense, aspect, subject, negation	Tense/Aspect: Future, Perfect, Imperfect, Past, Habitual; Negative; Person: 1, 2, 3; Number: sg, du, pc, pl; Clusivity: inc, exc.; Gender: SIB, MASC, FEM (see paradigms Walsh 1976: 326ff.)	L	"A verb root selects a nuber of the auxiliaries" (W 202). Also the inflection that an auxiliary uses (including syncretisms) is idiosyncratic (see appendix on W pp. 326ff.)	pos.	Can be stressed see pumá-dap-nu-ŋime-puru (use.hands.1inc.IRR-stop-FUT-PAUC.F-go.impf 'We will stop fighting.' (M 369)	yes	"Every verb must have an auxiliary" (W 202) "Auxiliaries which can occur without a verb root can be assigned a lexical meaning and this meaning to some extent reflects their meaning when in combiation with a verb root." (W 212)
-3.1	P/N/G	Direct object, benefactive.	Direct Object pronouns: ŋi (1sg) nyi (2sg) Ø (3sg) nyi (1.du.inc.M/F, 1pauc.inc.M/F) ŋanku (1du.sibling/exclusive.M/F, 1pauc.exc.M/F) nanku (2du.M/F/sibling, 2pauc.M/F) wunku/nku (3du.M/F/sibling, 3pauc.M/F) ŋan (1pauc.sibling, 1pl) nan (2pauc.sibling, 2pl) wun/n (3pauc.sibling, 3pl) (205-206). Benefactive pronouns: ŋa (1sg) mba (2sg) na (3sg.M) ŋe (3sg.F) nye (1du/pauc.inc) ŋaru (1du/pauc/exc) naru (2du/pauc) wiru/ru(3du/pauc.FM) ŋara (1pauc/pl.sibling) nara (2pauc.SIB) wira/ra (3pauc.sibling, 3pl) (W 208-209)	U	"The bound direct object pronominal forms are invariant (in any one tense) so that their description is considerably less complex than that of the subject pronouns." (W 205). No conditioning of benefactives except that they are mutually exclusive with objects (W 207).	pos.	Can be stressed see param-ŋí-wa-ninda pierce.3pl.nonFUT-1sg.OBJ-hit-du.M 'Two people are looking at each other.' (M 369)	no	Only in transitive clauses and with intransitive subject plurals. Benefactives can occur also in intransitive verbs, middle verbs (W 207), cannot occur when direct objects appear. (W 207).

MURRINH-PATHA (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-3.2	N	Subject	Subject number: ka (dual non-future in some stems) (N2010: 332); mutually exclusive with object (N 2010: 333). Stems with this suffix are e.g. 'say/do' (W 356), without e.g. 'become' (W 357)	L	Only present in the majority of auxiliaries (as such, lexically conditioned), elides when there is an object in P-3. (N 2010: 332-333)	pos.	Can be stressed, see M 368 ŋuddam-ká-wu-nime impel.reciprocal.1pl.non-future-pauc.subj-return-pauc.F 'we (paucal, masc) returned' (M 368)	no	
-3.3	N	Subject	ninda (dual); mutually exclusive with object (N2010: 334), if object is placed in this position, then ninda appears in P3 (N2010: 334).	G	Dual number is marked here when there is no object (transitivity-conditioned) (N 2010: 334-335)	pos.	Can be stressed, see dam-nindá-wa (pierce.3sg.nonFUT-du.M-spear 'two men speared him.' (M 369)	no	
-2	V	Reciprocal	nu (reciprocal) (see example W on p. 387)	X		pos.		no	
-1	LEX/V	Incorporate, applicative	No listing of incorporates that can occur with verbs in Walsh (1976). Compare examples in Nordlinger 2010: ma (applicative) (N2010: 326) rlarl (drop) (N2010: 327)	X		pos.		no	
0	LEX/ADV	Root, iterativity	Partial or full reduplication of the root possible to indicate iterativity (W 240ff.)	X		pos.	Can be stressed , see dem-ŋi-madárur pierce.reciprocal.3sg.nonfut-1sg.Object-anger ''m angry.' (M 368)	no	"Most verbs have a verb root although a verb may consist of a free standing auxiliary" (W 202)
1	A/T	Future, Perfective, Imperfective	nu (Future) m/n (Perfective) da (Imperfective) (W 214)	U	No conditioning found (W 214-215)	no		no	

MURRINH-PATHA (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
2	ADV	Incorporate (particles)	No listing of incorporated adverbs in Walsh (1976). Compare examples in Nordlinger 2010: deyida (in turn) gathu (towards) (N2010: 327)	X		no		no	
3	N	Subject, object, benefactive	ninda (1.du) ŋime (pauc.F) Ø (pauc.sibling) neme/name (pauc.M) ŋinda (du.F) (W 205-206)	G	"Whether the number marker is interpreted as ecnoding subject or object number depends on the form of the classifier stem [i.e. auxiliary] and the object marker." However, it does not seem to be determined by the lexical property of the auxiliary, but by the features it encodes. "The paucal marker [in P 3] refers to the subject [in P-4, encoding dual], unless there is a non-singular object [in P-3], in which case it refers to the object" (N 2010: 336)	no		no	
4	ADV	Auxiliaries, intransitive (marks pluractionality)	Frequently used: sit, habitual, stand, move (239). Used for marking pluractionality (NC2012: 9)	X	Suffix to a verb form to qualify the action of the main verb (W 239). "There is not a great deal of difference between the qualification given to a verb by HAB and STAND but particular verbs appear to prefer one over the other. No principled motivation has been discovered for this preference." (W 239)	no	Does not seem to bear stress of its own, see examples 28-29 in (M 369)	no	

NAVAJO (LCI sample); Young & Morgan & Midgette ("YMM") (1992); McDonough ("M") (2003) for establishing prominence

Prominence: phonemic contrast (having multiple contrasts, here: vowel constrast) yes: nasality + all tones no: high/low-tone, no nasality
"The last syllable in the word, the verb stem, is a content morpheme.(...) This syllable is prominent on many levels" (M 108) "The prominence seems to be governed by the inherent asymmetries in the phonotactics of each domain and the category of a morpheme." (M109)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-10	P/N	Oblique/ indirect object	sh/shí/shi (1sg) ni/ní/n (3s) bi/bí/b (3.yi/yí/y.PROX) yi/y´/y (3.OBV) ha/há/h (3) 'a/'á (3.non-specific) ha/ho/hw (Space/area) nihi/nihí/nih (1/2.du/pl) ádi/ádí/á/ád (self/own) 'ahi/'ałhi/ahí/'aLhí/'ah/'aLh (each other/each other's) (YMM 846)	U	No conditioning found except phonological (YMM 846)	no	Does not have all contrasts (YMM 846)	no	
-9	ADV	Incorporated postposition	Various postpositions co-occurring with P-10-pronominals. E.g. á (for, on behalf of) aa (to, about, concerning, off, by) cho' (support, help) ą́ą́'/ání (uncovering (as i removing dirt or leaves from a buried object)) (YMM 846) í/é (against, joining, reachign contacting, overtaking) (YMM 847) ya (under, beneath) (YMM 848)	X		pos.	Can have all contrasts (YMM 846-848)	no	
-8	LEX/ADV	Adverbial and thematic prefixes	Various tehmatic and adverbial prefixes, e.g. á (thematic with verbs of making, dig, awareness) cha (darkness) de (over, in excess of) (YMM 848) ka (chronically ill, invalid) (ntsi (mental state, apprehension, worry) (YMM 849) wó'ą/wó'ąą (over an edge, ito a ditch) dį́ (a group of four, a foursome) yá (talk) ni (a pair, a couple, two, by two's) (YMM 850)	X		pos.	Can have all contrasts (cf. 848-850)	no	
-7	ADV	Iterative	ná/né/ní/ń (Iterative) (YMM 851)	X		no	Only high tone (851)	no	
-6	N	Distributive, subject/ object	da/de/daa (distributive plural) (YMM 850)	U	No conditioning found except phonological (YMM 851)	no	Only low tone (YMM 850)	no	
-5	P/N	Object	shi/sh (1sg) ni/n (3s) bi/b (3..PROX) yi/i/yí/í (3.OBV) ha/ho/hw (3/Space/area) nihi/nihí/nih (1/2.du/pl) 'a/'e/'i/'o (3.non-specific) (á)di/(á)dí/(á)d (self/own) 'ahi/'ah (each other) (YMM 851)	U	No conditioning found except phonological (YMM 851)	no	Only high/low tone. (YMM 851)	no	
-4	V/P	Subject, passive	ji/dzi/zh/sh/z/i (3.polite/impersonal) ha/ho/hw (space/area/impersonal things 'a/'e/'i/'o/' (3.non-specific) 'di (agentive passive) (YMM 851)	U	No conditioning found except phonological (YMM 851)	no	Only low tone (YMM 851)	no	

NAVAJO (cont.)

Pos.	Catego-ries	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-3	LEX/ADV	Thematic/ adverbial	Different thematic/adverbial prefixes, e.g. di (occurs as a coponent of verb bases that involve movement of the arms or legs) di (appears as a component of some verb base that involve an elongated object) (YMM 850) hi/he/ha/hee/haa/ha/yi/i (seri-ative) si (occurs as a component of a verb meaning "kill one object") (YMM 852) ni (occurs as a component of verb bases that relate to the mind and to mental processes) ni (terminative) yi/yii/i/i/qo/o (transitional) (YMM 853)	X		no	Only high and low tone (550-553)	no	
-2	A/M/P	Subject, op-tative, sev-eral aspects/ aktionsarten	Different conjugations for subject depend-ing on aspect, e.g. ni-imperfective: nish (1sg) ní (2sg) (y)í (3sg) nii(d) (1du/pl) no(h) (2du/pl) aa (3pl) í (3.nonspecific) jí (3.nonpersonal) si-perfective: sé (1sg) síní (2sg) (yi)s (3sg) sii(d) (1pl) soo (2pl) aaz (3pl) iz (3.nonspecific) jiz (3.impersonal) (YMM 907) See. paradigms (YMM 907-921), also in combination with other af-fixes (P-8 – P-3)	L	The choice of con-jugation sets is lexically assigned. For example, the perfective stem '*mal'* can choose between a si-/Ø/yi-Perfective paradigm (YMM 395), whereas the root láá' (gather object together, collect object) can choose between si or Ø. (YMM 367). Lexically condi-tioned adverbial prefixes also con-dition conjugation pattern, as listed in YYM (867).	no	Only high/low/falling tone, no nasality. (907-921)	yes	Yound and Morgan analyze P-2 as two positions with aspect (preceding) and person (fol-lowing) slot. However, even forms are not always distin-guishable and contain both as-pectual and subject infor-mation in the paradigm. With the two-slots analysis in Young and Morgan zero forms appear, however, they suppose peg elements in the case P-2 is not filled, which can be in turn analyzed as part of P-2 paradigms. Phonologi-cally P-2 can be analyzed as obligatory (See paradigms on pp. YMM 907-921)
-1	V	intransitive, transitive, mediopas-sive, passive, reciprocal, causative; can be lexi-cal or deriva-tional	Ø (appears in 41 % of verb bases, active and neuter, transitive and intransitive) ł (appears in 28 % of the verb bases, active and neuter, trasitive and intransitive. Caus-ative. '/d (predominantly pasive, medi-opassive, reflexive or reciprocal deriva-tives, appears when first person plural) l (detransitivization of ł-valency) (YMM 883-885)	X		(no)	Consonantal, never appears at the end of the word.	no	

NAVAJO (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
0	LEX/A/T/M	Stem rhyme alternates for different aspectual/aktionsart, mood, tense.	Different stem sets for different verbs, and different alternations. E.g. for 'to smoke/burn': lííd (momentaneous imperfective/optative, neuter perfective) lid (momentaneous perfective, durative imperfective/optative/perfective, neuter imperfective) li' (repetitive imperfective, momentaneous/durative iterative) lił (Future) (YMM 370). Cf. with 'be wide broad: teel (neuter absolutiv, momentaneous perfective) téé (comparative) tił (iterative, future, progressive) teeł (momentaneous imperfective, optative) (YMM 498)	L	The alternations are somehow systematic, although not a fully productive system (M 47). The different sets of alternations as well as the form of alternations are idiosyncratic, see the Analytical Lexicon listing all the stem forms (YMM 1-798).	pos.	All contrasts (high, low, falling, nasal) (See verb stem index, YMM 1-798).	yes	
1	ADV	Several nominalizing, interrogative, adverbial enclitics (936-943). The ones that are written independently (i.e. are conisidered words) are excluded from this slot: (lágo, łeh, doo(leeł), ńt'éé.)	Examples: í (nominalization) (936) go (subordination) (937) dę́ę́' (from) (938) ísh (interrogative) (940) ni' (marks something from the past that the speaker remembers) (941)	X		pos.	All contrasts (high, low, falling, nasal) (See list of enclitics 936-943)	no	

NORTHERN POMO (UCI sample); O'Connor (1987)

Prominence: stress
"The root typically carries the accent". (20)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-2	ADV	Incorporation	Object, direction or goal, e.g. čaw (house) ʔuwi:l (up) lum (nettle) (32)	X		no	Never stressed; "do not change the stress pattern of the verb" (32)	no	
-1	V	Instrumental	ma (with foot) (17) di (with hands) ča (by sitting) phi (with a busting tool) da (with hand) mi (outward) ba (with words) pha (striking with hand) (18)	X		no	Not stressed because not a stem morpheme.	no	
0	LEX	Root	Usually has the form CV(:) (20)	X		yes	"The root typically carries the accent" (20).	yes	
1	LEX/ADV	Inherent aspect/manner	l (thematic.durative) c' (thematic.with pressure) (21) y (thematic.perfective/completive) (22) m (towards ro onto or on a surface) (28)	X		no	Not stressed because not a stem morpheme.	no	
2	ADV	Multiple event, directional	ta (multiple event) (53); various directional suffixes, e.g. (a)l (motion forward, hand over hand, or end over end; reeating cycles along a linear vector) (25) ew (start on or change trajectory" (27) mulu (around) (27);	X		no	Not stressed because not a stem morpheme.	no	
3	ADV	Inherent Aspect: continuative, semelfactive, progressive	m (continuative) (29) č/ʔ/či (semelfactive) (30) (a)d/n (progressive) (31)	X		no	Not stressed because not a stem morpheme.	no	
4	V	Causative, reflexive, passive	ka (causative) iʔ (reflexive) moʔ (reciprocal) ya (passive) (33)	X		no	Not stressed because not a stem morpheme.	no	

NORTHERN POMO (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
5	ADV/A	Progressive (inflectional) (also future /prospective reading), 'almost'	s'u (almost took place) (40-41) (a)d (Progressive/Future) (43-45)	U	No conditioning except phonological (43-46) "All verbs with just the progressive suffix -ad of course allow the progressive interpretation. Many also allow the prospective interpretation. However some seem to require the addition of the semelfactive to sanction the prospective interpretation. Many verbs have inherent semelfactive aspect, and these verbs all allow the prospective interpretation when the progressive morpheme is added.". However, O'Connor does not give examples for when this progressive affix is not used prospectively (the other two examples seem to use the derivational prospective morph in P3) (see examples on p. 44).	no	Not stressed because not a stem morpheme.	no	
6	T/M/PO	Present, remote past, past emphasis, perfect, ability, possibility, imperative, evidentials	e/Ø/o/u/a (present) (39-40) thi (remote past) (42) mi (past emphasis) (43) 46) do (speaker hears reports from ohers of occurrence of event denoted by host V) (290) nhe ("Speaker hears sounds of activity denoted by host V) (289) na (evidential) (46-47) nha (negation) (see example 148 on p. 48) malo (ability) (48) wa (possibility) (48) ya (hortative) am (imperative) (49)	U	Present: no conditioning except phonological (40); remote past: no conditioning except phonological (42). Past emphasis: no conditioning found (42-43); Ability: no coditioning found (48); Possibility: no conditioning found (48); Imperative: no conditioning found (49). Evidentials: No conditioning found (289-290): Negative: no information found.	no	Not stressed because not a stem morpheme.	no	
7	A/T	Past perfective	y/ye (perfective) (40) Appears after the Evidentials (47)	U	Perfect: no conditioning except phonological (40)	no	Not stressed because not a stem morpheme.	no	
8	A/T/M	Future	khémna (future:prediction) (45) khéna (intention) (45)	U	Future: no conditioning found (46);	yes	"It appears to be independent of the main verb, in that it bears stress."	no	

NUNGGUBUYU (LCI sample); Heath (1974)

Prominence: stress
"Long vowels attract high intonation and stress, especially when adjoining syllables have short vowels" (32) " When the last few syllables of a word have short vowels, the penultimate attracts high pitch and a little stress, and a pattern of alternating high-pitched, stressed vowels in even-numbered syllables (right-to-left) may result" (32) "In multisyllabic stems (especially nouns, the stem may be organised into "foot" units of two to three syllables based on apparent reduplicative segments, with each such unit having high pitch and some stress on the initial syllable." (32)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-4	A/T/M/P/N/G	Subject/object, tense/mood/aspect	Various, numerous allomorphs that mark subject/object (person, number, clusivity, gender) and TMA (Set A/B) together, e.g. nga (1sg.A) ngan (1sg.B) ngi:ni (1IntMtr.A) ba (2sg.B) nimbingi (2.F.du.B) etc. (348) Ø (Gender ANA/MANA acts on Gender ANA) (357) (348-361). Person: 1, 2, 3; Clusivity: inc, exc; Number sg, du, pl; Gender: ANA, WARA, NgARA	G	Allomorph sets in P-4 and suffixes in P3 are co-dependent categories, only in combination they render Past Actual, Potential, Present, Future, Evitative, Negative, Continuous, Punctual (338-339)	pos.	Can fill out every second preceding syllable. (32)	(yes)	Heath does not mention explicitly that these prefixes are obligatory, but the description of the grammar suggests so. " The term pronominal prefix designates a class of prefixes (many of which cacn be decomposed into two or more component morphemes in a phonologically abstract analysis) which are used with erbs, and with (...) adjectival nouns (..) when they are in predicative form." (347). Only one instance is zero marked, the 'equipollent' constellation (357-359), which is zero syllabic. All other allomorphs can be mono or multisyllabic.
-3	ADV/V	Comitative, benefactive, multiplicity	ag/wa:G (benefactive (377) anyji (Comitative) (381) ngaran/ngaraG/wara/waraG/hagara/waragara/walgara (multiple prefixes, determines the mutliplicity of agents) (382)	X		pos.	Can fill out every second preceding syllable.	no	
-2	ADV	Repetition/ prologation	Reduplication of some compound stems (342)	X		pos.	Can fill out every second preceding syllable.	no	
-1	LEX/ADV	Compound stems	Compound stems for some verbs, e.g. with lharma (hunt) wadji 'freshwater animals' walhaga 'go looking'; with na (burn) wu- 'make huge campfire'. Many of thee compunds are lexical (478)	X		pos.	Can fill out every second preceding syllable.	no	

NUNGGUBUYU (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
0.1	LEX/ADV	Root, repetition/ prologation, reduplication	Internal reduplication or prefixal reduplication (342)	X		pos.	Can fill out every second preceding syllable.	yes	
0.2	LEX/T/M/A	Root, formatives that are dependent on T/M/A	a/a:/i/i:/u/u: (formative) (408-411)	L	Whether long or short vowel is determined by TMA categories and verb class. For example, Verbs from class A2 use long a: in Nonpast, whereas verbs from class A3 use short vowel (408).	pos.	Can be penulitmate position and long in ultimate position, ergo stressed.	yes	Every stem has a final vowel.
1	A/T/M	Part of formative for position 2 TMA 'Augment'	n (409) ya (409) nga (410) ma (410); 'Inchoative' (395) ra (410) yi: (411)	L	Unpredictable insertion of these formatives with suffixes in P 3 dependent on class (408-411)	pos.	Can be in penultimate position, ergo stressed.	no	
2	ADV/V	Inchoative, causative, reciprocal, reflexive	wi/dhi (Inchoative) (395) Ø (Reflexive) nyji (Reciprocal) jga (Causative) (408-411)	X		pos.	Can be in penultimate position, ergo stressed.	no	
3	A/T/M	Past 1/2, Nonpast 1/2/3, Evitative. Rendered categories in combination with P-4: Past Actual, Potential, Present, Future, Evitative, Negative, Continuous, Punctual	ny (past/nonpast 1) ngi (past2, nonpast3) ng (nonpast 1, past 1) na (nonpast 2) ni (nonpast 2, past 2) Ø (Nonpast 3, Past 2, Nonpast 2) ngun/ngan/nyji (evitative) di (Past 2) ji: (Nonpast 2) jan/n (Evitative) ra (Nonpast 2) y (Past 2) (408-411).	L	Different allomorphs for different TMA, depending on class (407-422)	no	All allomorphs, if they constitute a syllable, are short, and therefore not stressed.	no	Different zero allomorphs for a variety of categories. (408-411)

NYAMWEZI (GCI sample); Maganga & Schadeberg (1992)

Prominence: high tone
"The general rule is that any underlying high tone is realized not on the vowel to which it lexically belongs but one mora further to the right" (42) Low tone seems to be unmarked, for example, verbs have either all syllables low tone or one marked. (98). Most of the affixes have realized low tones (se charts on pp. 102, 103, 104)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-6	ADV/T	Sequential, associative	na (sequential past) (103) na ('and, with') (109)	U	No conditioning found (109).	no	No inherent tone (109).	no	
-5	P/N/G	Subject	na* (1sg) u* (2sg) tu* (1pl) mu* (2pl) a/u (class 1) ßa* (class 2) gu* (class 3) yi (class 4) li* (class 5) ga* (class 6) ki* (class 7) shi* (class 8) yi (class 9) zi* (class 10) lu* (class 11) ka* (class 12) tu* (class 13) ßu* (class 14) ku* (class 15) ha (class 16) ku (class 17) mu class 18) (102). *These listed forms bear a high tone (non-low subject forms in chart on p. 102), but since the tone gets realized on the following syllable, it is never realized on the subject concords.	G	The choice of third person morphemes depends on the noun class of the subject referent (101)	no	Although some morphs bear inherent high tone, the tone does not get realized on the morph itself. (102)	yes	"All verb forms with the exception of Imperatives must have a subject concord (101)
-4	PO	Negative	ká/ka (negative) (106-107)	U	No conditioning found except phonological. Tone depends on the Subject position (P5) (107)	pos.	Has inherent tone. The inherent tone can be blocked in some contexts (107)	no	
-3	T/M/A	Future, habitual, progressive, narrative, consecutive, inchoative, immediate past, recent past, distant future, habitual hortative	ku (aorist) lii (progressive) ka (narrative, stays low) ú (consecutive) úyuú (inchoative 1) ayúú (inchoative 2) a (past) (103) laa (future) (105)	U	Although the markers are used as formatives for T/M/A constructions, they have their own meaning which does not change in combination with other markers (105)	pos.	Can bear high tone if the stem has a lexical tone (98-101)	no	This position is not filled in sequential past, perfective, optative, hortative and imperative.

NYAMWEZI (cont.)

Pos.	Catego-ries	Functions	Morphs/Features	Cond.	Explana-tion Cond.	Prom.	Explanation Prom.	Obl.	Explana-tion Obl.
-2	ADV	Itive	ka (go and/going to) (107-108)	X		pos.	Can bear tone either inherent or when morphs from preceding position have inherent high tone (e.g. wa-á-gwiilé) (105), see chart on pp. 103-104)	no	
-1	P/N/G	Object	ni (1sg) ku (2sg) tu* (1pl) mu* (2pl) mu (class 1) i* (refl) (102) ßa* (class 2) gu (class 3) yi* (class 4) li (class 5) ga* (class 6) ki (class 7) shi* (class 8) yi (class 9) zi* (class 10) lu (class 11) ka (class 12) tu* (class 13) ßu (class 14) ku (class 15) ha (class 16) ku (class 17) mu (class 18) (102). *These listed forms bear a high tone (in the description: plural participants, 101), but since the tone gets realized on the following syllable, it is never realized on the object	G	The choice of third person morphemes depends on the noun class of the subject referent (101)	pos.	Can bear tone when morphs from preceding position have inherent high tone (e.g. u-ní-bonágé) (108)	no	
0	LEX	Root	CV(VN)C(V(VN)C) syllable shape, high or low tones. (98) The stems themselves have two tone classes: one which is low throughout and one which has just one high tone (underlyingly) on its first vocalic segment. On the surface, this underlying high tone is shifted to the right and appears on the next one, two or even three moras. The almost 700 verbs in our Vocabulary are nearly equally divided between the two tone classes (98).	X		pos.	Can bear high tone if the stem has a lexical tone (98-101).	yes	
1	A/T/M	Formative	ag (habitual/recent past/habitual hortative, past intentional, imperative) (103-105)	G	Meaning depends on the construction (105)	pos.	Can bear high tone if the stem has a lexical tone (98-101).	no	Only found in few T/M/A's (103-104)
2	A/T/M/N	Past, non-past, imperative, imperfective, future, optative, imperative plural	Final vowel a (past/non-past) á (remote past/imperative) ilé (perfective) (e)é (future/optative/hortative/past intentional, imperfective) (103-104) i (imperative and 1st person plural hortative) (109). The imperative plural is listed as a separate position but replaces the final vowel. As such, it is written here in the same position with the final vowels (see paradigms on p. 133)	G	Meaning depends on the construction (103-105)	pos.	Can bear high tone if the stem has a lexical tone (e.g. ukáßon-ág-é, p. 108). –i (imperative plural) can bear high tone if the final vowel that it replaces has an inherent high tone (109).	yes	Found in all forms (103-104)
3	ADV	Locative enclitics	h(o)ó k(o)ó m(o)ó (locative enclitics) (143)	X		pos.	Locative enclitics can bear high tone, e.g. akaja-moó (109)	no	

ONEIDA (LCI sample); Abott (2000)

Prominence: stress
"The general rule is to count back two syllables (two vowels) from the end of the word and in so doing skip any stem joiners or epenthetic vowels before the aspect suffix" (9)

Pos.	Catego-ries	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-10	ADV/PO	Partitive, coin-cident, contras-tive, negative	n(i) (partitive) tsh(i) th(i) (contrastive) teʔ (nega-tive) (11, 13)	U	Negative does not occur with the punctual aspect suffix (20). "The par-ticular verb stem and the aspect suf-fix determine whether the locative prefixes (cislocative, and transloca-tive) indicate location or direction" (20). Vowel elision phonologically conditioned (13).	no	Does not seem to attract stress (Too far away from the penultimate sylla-ble)	no	Negative not required (38)
-9	ADV	Translocative	y(e) (translocative) (11, 13)	X		no	Does not seem to attract stress (Too far away from the penultimate sylla-ble)	no	
-8	A	Aorist, see slot -6	waʔ/u/aʔ((aorist; fac-tual) (11, 13)	G	Sub-aspect with aorist morpheme in P-4. Allomorphy of the aorist due to different combinations with preced-ing and following affixes. For exam-ple, when aorist appears in all three positions (-8, -6,-4), then aorist here is aʔ; e.g. y-aʔ-t-u-s-a (translocative-aorist-dual-iterative) This position does not show aorist for iterative, cislocative (see table on p. 13)	no	Does not seem to attract stress (Too far away from the penultimate sylla-ble)	no	Aorist not required (38)
-7	ADV	Dualic, recipro-cal	teʔ (11, 13)	X		no	Does not seem to attract stress (Too far away from the penultimate sylla-ble)	no	
-6	A/T	Future, indefi-nite, aorist	ʌ (future) a/uu (indefi-nite) a/u (aorist) (11,13)	G	Allomorphy of the aorist and indefi-nite due to different combinations with preceding and following affixes. This position does not show aorist for iterative, cislocative (see table on p. 13). These cooccur with punctual as-pect suffix.	no	Does not seem to attract stress (Too far away from the penultimate sylla-ble)	no	Future/Indefinite not required (38)

ONEIDA (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-5	ADV	Iterative, cislocative	s (iterative) t (cislocative) (11,13)	X		no	Does not seem to attract stress (Too far away from the penultimate syllable)	no	
-4	A/T	Aorist, indefinite	a (aorist/indefinite) (13)	G	Appears only in combinations with some adverbial elements, e.g. n-u-t-a (partitive-aorist-iterative-aorist) (13)	pos.	Stress in some inflected words, like wá-hl-eht-eʔ (aorist-3sg-go-punctual) (15)	no	Appears only in combinations with some adverbial elements, e.g. n-u-t-a (partitive-aorist-iterative-aorist) (13)
-3	P/N/G	Object, subject	Different scenarios where subjects act on objects and vice-versa different for different stems; e.g. a-stems: skw (1sg>2sg) wak (3.N > 1sg) yetshiy (3non.sg > 2); c-stems: sk (1sg>2sg) wak (3.N > 1sg) yetshi (3non.sg > 2) (22-31); person: 1,2,3; number: sg, du, pl; clusivity: inc, exc; gender: masc, fem1, fem2, neu.	L	"In general the verb stem determines whether subjective, objective, or transitive pronominal prefixes are used" (21) ... "Most dynamic verb stems require subjective prefixes, except when the stative suffix is added, while others require objective prefixes (38)	pos.	Stress in some inflected words, like ye-há-ha-s (translocative-3sg-carry-serial) (16)	yes	No zero forms (cf. pp. 22 - 31)
-2	ADV/V	Reflexive	at(ʌ/e)/an/al (reflexive) (32)	X		no	Does not seem to attract stress (see pp. 32-33)	no	
-1	LEX	Incorporated Noun	noun roots, nominalized verb stems, empty roots (semantically empty morphemes required by certain verbs when no specific noun is incorporated) (33)	X		pos.	Must be rare since the root usually attracts stress. One example is t-ka-núhs-ot-eʔ CIS-PRO-house-stand-STAT 'there is a house standing' (17).	no	

ONEIDA (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
0	LEX	Root	lexical component (34)	X		pos.	Very commonly attracts stress (see examples, e.g. on p. 39 s-a-shako-téwaht-eʔ (iterative-aorist-3>3-miss-punctual)	yes	Can be monosegmental (k; e) or multisyllabic (nuhwelatu) (34)
1	ADV/V	Dative, instrumental, stative, distributive, dislocative	hslu/hu/nyu/u (different distributives) (a)ʔ (non-stative) ht/ʔt/st/t/hkw (different instrumentals) (35) ʔse/ʔs/ni/ʔs/ʔseni/ʔs/ʌ(ni) (Dative; alternating for serial/punctual/stative suffixes). The choice is controlled by the verb (37). Dative allows a pronominal prefix to refer to a patient or a beneficiary (37); hsy/kw/kw (undoer) h/ʔn/hn/hsl/aʔn/ahn (dislocative) (37).	X		no	Does not seem to attract stress (see pp. 34-37)	no	
2	A/M	Seriative, punctual, imperative, stative	heʔ (serial) hah (serial past) keʔ/ʔ (punctual) Ø/u (stative) imperative) Ø/(ha)k (imperative) hne (stative past) (42, 43, 45)	L	"A large class of verb stems that are inherently dynamic rather than stative inflect for four suffixes: serial, punctual, imperative, and stative" (42). "Determining which verb stems will express ongoing current action with the serial and which with the stative is not entirely straightforward" (42). Abbott 2000 gives an example at at the bottom of p. 45 with an imperative suffix which is not commented on. "The serial, punctual and stative suffixes can all inflect for past and future tense" (43).	pos.	Stative can be accented (ú, p. 44).	no	
3	A/T	Progressive, past	tye/atye/hatye (progressive) (43) (w)e(ʔ) (past) (44-45)	G	"As with motion verbs, verbs with Inflect with stative/serial/punctual (43)	pos.	Progressive can be accented (hátye, p. 44)	no	

PILAGÁ (LCI sample); Vidal (2001); Klein (1974) for prominence

Prominence: stress
Stress falls on the last syllable of the word (Klein 1974: 30) but Vidal (2001) says stress is "lexical and unpredictable" (70). " The actual stress bearing-syllable will then be that which is phonetically second or third" (70). Certain suffixes always carry stress. These are aspect, object (number) agreement and directional markers) (72). "When words have two or more suffixes, the root retains its stress; whereas primary stress is assigned to the suffix, which is the rightmost stress bearing unit".

Pos.	Catego-ries	Func-tions	Morphs/Fea-tures	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-4	PO	Negation	sa (negation) (283)	U	No conditioning mentioned	no	"Nominal and verbal prefixes are always unstressed." (71)	no	
-3	P/N	Object	yi/ñi (1sg) an (2sg) Ø (3sg) qom'i (1pl) Ø (3pl) am'i (2pl) (143)	L	"Allomorphs for 1st person singuar are not phonologically triggered, but are lexically assigned."	no	"Nominal and verbal prefixes are always unstressed." (71)	no	"Unlike subject prefixes which can co-occur with a pronoun or full NP, object prefixes are mutually exclu-sive with full pronouns and lexical NPs." (142)
-2	P	Indefi-nite sub-ject	qo (indefinite subject) (146)	U	Co-occurs oblgatorily with a third per-son subject prefix from either set A or B (although it doesn't change its form or meaning) (146). Therefore, the pre-fix itself is non-conditioned.	no	"Nominal and verbal prefixes are always unstressed." (71)	no	
-1	P	Subject	Set A: s (1) aw/o (2) d/t/i/yi/h/Ø/w (3) Set B: ñ (1) an (2) n (3) (136)	L	Lexically conditioned, neither mor-phology nor phonology can predict the classes (91). Verbs fall into three categories: the ones that have set A, others that have set B and the ones that alternate between A and B (here, the conditioning is semantic) (177). Otherwise, "the assignment of a verb prefix calss is synchronically lexical-ized" (178).	no	"Nominal and verbal prefixes are always unstressed." (71)	yes	Vidal (135) does not explicitly say that each verb has a prefix set, but " pronominal prefixes constitute the only case marking device, since the language lacks case markers on nouns and free pronouns, and lacks adpositions for nominals." (136). In addition the author notes that " Pil-agá verbs can be classified into three groups according to the way subejct marking is distributed: a) those that can take Set A prefixes only; b) those that can take Set B prefixes only; and c) those that can take ei-ther Set A or Set B prefixes." (137). "In the third group, there is a fairly regular semantic contrast between the Set A marked and Set B marked verb forms (...). Though this schema accounts for the vast majority of the prefix choices on Pilagá verbs, the assignment of a particular case to a verb is still lexicalized." (137)

PILAGÁ (cont.)

Pos.	Catego-ries	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
0.1	LEX	Root	Lexical root (157).	X		pos.	Can be the last syllable of the word with non-progressive suffixed (see examples 42b-h on p. 272) or Ø-marked non-progressives (see examples 43 on p. 273).	yes	
0.2	N	Subject	i (2pl) d (1pl) (151) (infix versions of P2)	L	1st and 2nd plural can be infixed to the root (see examples 151)	pos.	Can be the onset of the last syllable of the word, such as in na-nom<d>a setA3-know<pl> 'They know' (151)	no	
1	LEX/ADV	Derivations and aga, which is somehow thematic, p. 86.	aʕan (transitivizer) l'at (reflexive) 'at (reciprocal) (166)	X		pos.	Can be the last syllable in the word, when aspect markers is zero or switches position and comes before it. See example 17 a) on p. 168: ñi-tawa-n'-at set.B.1-help-ASP.non.PROGR-RECIPR 'We help each other' or 20 on p. 171.	no	
2	N	Subject plural (both transitive as well intransitive subjects, p. 161)	(a)q/soq/ʕa/sa (1pl) i/e/q(a)e (2pl) d/y (3pl) (150, 152) 1pl and 2pl can be infixed in P0 (136/157)	L	"I conclude that [...] the distribution of subject plural allomorphs must be lexically specified"	pos.	With aq, the suffix can be the last syllable in the word, as it switches position with the directional marker in P5, see example f) on p. 155: ñ-atoʕo-g-aq setB.1-spt-DIR-PL 'We spit'. Vidal (73) mentions following stress bearing morpheme: l'at (reflexive). Also see 2nd plural attracts stress (161): an-qač-í-ñi setB.2-catch-2.pl-DIR.downwards (you all caught.') (161).	no	

PILAGÁ (cont.)

Pos.	Catego-ries	Func-tions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
3	A	Progres-sive, du-rative, habitual, non-pro-gressive, complet-ive, re-sultive	ta(k)(progressive) tape (pro-gressive 3pl) tapiñi/tapiyi (du-rative) (ta) pega (habitual) n/Ø (non-progressive) tayi/ñi/yi (completive) tañi/ta (resultive) (259)	L	Progressive allomorphs are determined by mor-phological (3pl) and phonological criteria (260). The allomorphy between tapiñi/tapiyi (durative/progressive/iterative) is lexically con-ditioned (263). "Some verbs indicate 'non-pro-gressive' aspect by no surface marker at all (272). "The reason why some verbs have gram-maticized to select one particular [completive] allomorph is unknown for the moment" Some verbs "will fail to take morphology to distin-guish both categories; such verbs are simply grammaticized to select either the resultative or the completive form (or zero) to indicate that the action is finished (274)	pos.	Can be the last sylla-ble in the word. Vidal (73) mentions the fol-lowing inherently stress bearing mor-pheme: péga 'non-progresisve/habitual aspect'	(yes)	Klein (1974) says one of the suffixes in this position is obligatory (127), but there are zero punctual (non-progressive for Vidal 2001) allomorphs that are lexically condi-tioned.
5	ADV	Direc-tion and emotion	ge' (away from the reference point) get (towards the refer-ence point) segem (upwards) ot (upwards; under) ñi (down-wards) som (downwards; e.g. downwards a water source) som (downwards, to the in-side) owe (inwards) wo (out-wards) ege (forward; in front of) eg'a (to/in a specific place) iyi (in straight line) sop (in cir-cles; with) lege (on/over) ta (out of ; to the other side) pe (along with; concurrent mo-tion) (231)	X		pos.	Can be the last sylla-ble in the word. Vidal (73) mentions fol-lowing inherently stress bearing mor-phemes: tá, pége	no	
6	N	Object/ intransi-tive sub-ject number	a (singular) to (paucal) lo (plu-ral) (165)	L	Applies to objects but also to intransitive sub-jects in some verbs. "The reason why [...] plural number of the subject is indicated through the object number markers, rather than by subject number markers, has no explanation to my knowledge. One possibility is that this plural marking system may be the residue of a no longer productive absolutive agreement system, or rather the opposite, it could be an emergent absolutive system, and for either reason some-times functons to indicate the plural number of the only argument of an intransitive verb" (165)	yes	These Suffixes are al-ways the last syllable of the word and carry always stress (72-73)	no	Object number suf-fixes are optional (162)

PIRAHÃ (UCI sample); Everett (1986)

Prominence: stress
"The rule for stress placement may be stated informally by saying that primary stress is placed on the heaviest of the final three syllables in the word. In the event that the heaviest syllable type occurring in a word has multiple tokens in that word, the most rightward token will be stressed.

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
0	LEX	Lexical root	Lexical Root	X		pos.	In the case that the stem is heavy and only light syllables follow.	yes	
1	ADV	Incorporation	e.g. op 'go' hoag 'come' ap 'turn' (301)	X		pos.	In the case that the incorporated stem is heavy and/or only light syllables follow.	no	
2	A	Durative, punctiliar	ab (durative) (294); occurs often with continuative in P7 (294) ap (punctiliar) (295). These aspects cannot co-occur with P3 (telicity), P4 (aspect) P13 (frustration) (290), which means that they can be considered as inflectional as the others.	U	No conditioning found (294)	no	Is not stressed, since the syllables are not heavy and they do not occur at the end of the word the syllable type VC is not possible, therefore, an affix must follow (311)	no	
3	A	Telic, atelic	áo (telic) (290), used in perfective aspect most often with P4 -b (290). ái (atelic) (291), used in imperfective aspect most often with P4 -p (291).	U	No conditioning found (290-291)	pos.	Can be stressed when both syllabic morphemes are heavy syllables	no	
4	A	Perfective, imperfective	b (perfective) (290) p (imperfective) (291) (used to make time reference with P3 aspect suffixes and P10 suffixes and P12 modal suffixes) (291).	U	No conditioning found (290-291)	(no)	Can form part of the preceding stressed syllable (P 3) they cannot occur as (final) codas; the syllable type VC is not possible, therefore, an affix must follow (311)	no	
5	M	Desiderative	sog (desiderative) (296).	U	No conditioning found (296).	no	No examples found where this morpheme is in final syllable position (it always is a light syllable, and a vowel always follows /g/).	no	

PIRAHÃ (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
6	PO/M	Negative, prohibitive	hiab (negative) (250) sah(ax)ái /saí (prohibitive) (249).	U	No conditioning found except phonological (249-250)	pos.	Stressed if these morphemes are final syllables and are not followed by heavy syllables.	no	
7	ADV	Continuative	xiig (continuative) (292).	X		pos.	Stressed if it is the final syllable and is not followed by heavy syllables	no	
8	ADV	Interrogative	(xó)xóí (interrogative) (237) hoaxái (interrogative, questions dealing with existence or possession) (237-238) xaoxái (lack of control over the response) (238).	U	No conditioning found (237-238)	pos.	There is no example where these morphs do not occur as last syllables of the word (238-239).	no	
9	ADV	Ingressive	hoag (beginning of a state) hói (beginning of an action) (292).	X		pos.	Stressed if these morphemes are final syllables and are not followed by heavy syllables.	no	
10	T	Proximate, remote	i (proximate) (293) a (remote) (293) (used to make tense reference, probably non-present?) (290-291).	U	No conditioning found (293)	pos.	Only in verbs where it is the last syllable of the stem and no other heavy syllables precede, like kob-á see-remote ('look!') (274). However, stress must be very unlikely since these suffixes are preceded by heavy syllabic aspectual suffixes or followed by other suffixes.	no	
11	ADV	Iterative	ta (iterative) (292).	X		pos.	If it is the final syllable and is not preceded by heavy syllables, like in kahá-pi-tá (leave-imperfective-proximate-iterative) 'already left' (227)	no	

PIRAHÃ (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
12	M	Certainty	áti (uncertain), occurs in imperatives (295) haí (relative certain) há (complete certain) (295).	U	No conditioning found (295)	pos.	If these morphemes are final syllables (and áti and há are not preceded by heavy syllables)	no	
13	ADV	Frustrative	ábagaí (frustrated initial, almost began to) (300) ábai (frustrated terminal, almost finished) (300).	X		yes	There are no examples where these morphemes are not in syllable-final position or where a heavy-syllabic morpheme follows.	no	
14	ADV	Intensive	baí (intensive) (299).	X		pos.	The only example where this morpheme is not in syllable-final position and is followed by a heavy-syllabic morpheme follows is when the second emphatic marker is attached: ti gíxai xog-i-baí-koí 1 2 want-epenthetic-intensifier-emphatic 'I really like you a lot.' (299)	no	
15	ADV	Emphatic	koí (emphatic) (299).	X		yes	There are no examples where these morphemes are not in syllable-final position.	no	
16	ADV	Nominalizer	so/ao (temporal nominalizer) (263-264) (i)sai~saí (conditional/valence-reducing nominalizer) (219, 264, 279) si (nominalizer, proper nouns) (279).	X		pos.	In the case that the light syllable morphemes are at the end of the word with no heavy syllable preceding or the heavy syllable morphemes are at the end of the word or are not followed by heavy syllables.	no	
17	ADV	Evidentiality	híai (hearsay) (298) xáagahá (observative) (298) sibiga (deductive) (297).	X		yes	There are no examples where these morphemes are not in syllable-final position or where a heavy-syllabic morpheme follows.	no	
18	ADV	Clause marker	taío (result; "and therefore") (297).	X		yes	The clause marker is always the last syllable of the word.	no	

SEMELAI (UCI sample); Kruspe (2004)

Prominence: stress
"In Semelai the domain of word stress is the final syllable and there is no secondary stress. Only phonological words bear stress. In the case of words bearing suffixes, stress shifts from the root to the suffix." (40)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-6	PO	Negation	daʔ (negation) (349)	U	No conditioning found (316).	yes	represents an own word, therefore stressed	no	
-5	M	Necessity, possibility, succeed	mstiʔ (have to, must, be obliged) (165) moh (want) (166) lən (desire) sot (can be able) (168) dapat (to get to) (!69)	U	mstiʔ, moh, lən (desire): No conditioning found (165-166) sot: no conditioning found. (168) dapat: No conditioning found (169)	yes	represents an own word, therefore stressed	no	
-4	A	Imminent aspect	ga (imminent aspect) (163). Occurs in several with several functions (163-164), used with future meaning.	U	No conditioning found (163)	no	Not stressed, because clitic	no	
-3	M/P/N	Agentive, irrealis agentive	ma (irrealis/potential agentive) (161) ʔəɲ (1minimal.familiar) kɒ (2.minimal.familiar) yɛ (1.minimal) ji (2.minimal) hɛ (1+2) ki (3sg) de (3pl) ko (3.unidentifiable) (171) (transitive actor, or (e)motional actor) (157)	U	Agentive proclitics: No conditioning found. Only appear in transitive verbs or verbs of actions; this does not make it grammatically conditioned per se, because there is no opposition and no allomorphs. (157). Irrealis/potential proclitic: no conditioning found. The irrealis agentive occurs, like the pronominal clitics, with transitive verbs as well with verbs of emotion (162), or inherent qualities of inanimate entities (162).	no	Not stressed, because clitic (62)	no	

SEMELAI (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-2	ADV/V	Imperfective intransitive verbalization, agentive intransitive, middle, causative, happenstance, comparative, intensive, collective	Reduplication of the onset and coda of the root (imperfective intransitive verbalization) (110) mN (Agentive intransitive derivation; it is used in most varieties when the verb is active and refers to some definite process, or as a marker of durative, habitual or causal aktionsart; performatives) (115, 116, 152) b(r) (middle voice, occurs thematically too, decausative, reciprocal, intensive stative, auto-causative, reflexive reciprocal) (117-122) br (collective; cirumfix with -an) (123) par (causative monosyllabic root deverbalizer and denominalizer) tar (causative monosyllabic root deverbalizer) p (causative bisyllabic root deverbalizer, denominalizer) (124, 127) t(r) (happenstance, inability) (140, 143) t/raʔ (comparative) (146).	X	The imperfective reading is a connotation of this derivation ("imperfective aspect associated with intransitive events.") (111) "Transitive verbs are associated with unitised perfective aspect, whereas intransitive verbs have inherently unbounded imperfective aspect" (111). However, this morpheme can occur in transitive clauses, rendering an iterative meaning (114).	no	Not stressed, because clitic (62)	no	
-1	ADV	Intensive	Light syllable reduplication (intensive) (149)	X		no	Not stressed, because it is not the last syllable of the stem.	no	
0.1	ADV/V	Root, imperfective intransitive derivation, causative	Sometimes the imperfective intransitive derivation is expressed by the root coda infixed: j<y>ʔoy make<intr> (113) r (causative bisyllabic root deverbalizer, denominalizer) (124). Sometimes the comparatives occurs as an infix: j<raʔ>leʔ 'to be shorter.' (147)	X		no	Infixes are not found in last syllabic position.	no	
0.2	LEX	Root	Usually (CV)CV	X		pos.	The final syllable is occupied by the stem.	yes	
1	ADV/V	Applicative, collective	an (collective; cirumfix with -br) (123) iʔ (applicative for motions and emotions) (135) (iterative) (138)	X		pos.	Suffixes are always stressed, since they are the last syllable of the word.	no	

SEMELAI (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
2.1	ADV/M	Discourse markers (clitic)	sɒn (speaker conclusion) (414) ɲa (clitic) ɲa (anaphoric) (418) pa (factual) (419)	X		no	These clitics do not take stress. ɲa only is stress in imperative contexts (P2). (417).	no	
2.2	M	Imperative (clitic)	cəʔ (emphatic; realis actuality, imperative) (411) ɲá (imperative.transitive) (417)	U	Free variation of clitics "The imperative is also distinguished from the delcarative by the optional presence of enclitic illocutionary particles. cəʔ: no conditioning found (411), ɲá: no conditioning found (417) "The two enclitics occur frequently and appear to be freely interchangeable. These are not solely imperative clitics; they also function as dicourse clitics" (332)	yes	both imperative markers are always stressed (417)	no	

SHEKO (LCI sample); Hellenthal (2010)

Prominence: tonal variability
Four level of height (1,2,3,4); 2 is unmarked (111).

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-1	P/N/G	Subject (Clitic)	n (1sg; syllabic) ha (2sg) há (3sg.M) yí (3.sgF) ń (1pl; syllabic) ítí (2pl) íʃì (3pl) (323, 429)	G	Everywhere except in realis, obvious and optative (289) See pp. 437/438 for an example where the clitic occupies the slot of the aspect. Not present in Subject focus constructions (436)	pos.	Three tones (429)	no	Main clauses which do not contain subject clitics are those where the subject is the focus. (436)
0	LEX/M	Root, factual, non-factual, causative, passive	Different Vowel+velar (319) Tonal alternations for mood. Only a small subset of verbs alternate for the presence or non-presence of a velar (320) Tone alternation for Factual (realis, obvious, viewpoint, implicative)/Non-factual (Irrealis, negative, imperative plural, optative) (298, 299) Can changed according to causative (374), passive (379)	L	"Three groups of tonal alternation on the basis of the tone of the verb stem (...) this correlation is not absolute and the groups are defined purely on the basis of tonal behavior" (However, in order to know which tone a verb has, one has to know the stem) (298) Non-velar stem is used: before the negative ara, before the same subject converb tə, as first member in verb-verb compound stems, in Realis forms if the subject clitic precedes it, in Irrealis forms if the Irrealis marker -m folows directly. (317) Tone varies according to mood (297)	yes	Four levels of height (114-115)	yes	Always present, usually monosyllabic: CV(V), CVC, CVVC. Disyllabic verb roots are relatively few in number, and trisyllabic verb roots have not been attested (81)
1	V	Causative, passive, middle	s (causative) (373) (can be infixed or put after the epenthetic vowel (374) t' Passive (379) (Can be also infixed) (380) n (384) (middle, assimilates to the adjacent consonant, 385) u (expletive vowel, used when a stop follows the stem. Usually present in derivations (373)	X		yes	Four levels; see alternation for syllabic middle according to assimilation to the stem tones (388)	no	

SHEKO (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
2	ADV/A	Imperfective, perfective, irrealis	ki (imperfective, 'exist, live, stay') (309) k'é (perfective, 'be left, remain') (310) a ('put', does not add meaning; all tones copied from the verb stem) (occurs with Irrealis, alone only in questions, 313) (312)	G	a' depends on several categories (Irrealis, Questions, can only be combined with Perfective (k'yá) (would not) (314)	yes	Four tones (assimilation included)	no	Element denoting aspect is optional (289)
3	P/N/G	Subject	n (1sg; syllabic) a (2sg) á (3sg.M) í (3.sgF) ń (1pl; syllabic) ítí (2pl) íʃì (3pl) (323)	G	In realis, obvious and optative the clitic follows the verb (289) See pp. 437/438 examples (24 a-b) where the clitic occupies a position before modal (P-4) but it is not clear if it represents the same position as aspect.	pos.	Three tones (429)	no	Only present in realis, obvious and optative (289).
4	M/PO	Imperative, jussive, optative, irrealis, negative, realis, obvious, viewpoint, implicative	Ø (Imperative, Jussive, Interrogative) s (Optative) m (Irrealis) ara (Negative) k (Realis) kn (Obvious) s (Viewpoint) a (Implicative) (297)	U	No conditioning found (297ff.)	no	All consonantal except in negative, which has tone Number 2 (297).	no	Several mood markers are zero (297)
5	M	Indirect/ direct stance	ə (indirect stance; certain distance between the speaker and the utterance) o (indirect stance, questions/vocative ya/a (direct stance) (absence of distance; makes the utterance more direct and less polite) (292). "a occurs only with the Viewpoint, Implicative; ya with obvious and imperative for children. (294).	G	Dependent on mood (297)	no	Only tone Number 2 (293)	no	"Stance markers are not obligatorily present, but they are common." (292)

SKOU (LCI sample); Donohue (2004)

Prominence: stress
"Stress is thus completely predictable, and is assigned to the first syllable in a simple word. In a word with proclitics, we find that stress remains on the first syllable of the root." (82)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-1	P/N/G	Subject (clitic)	nì (1sg) mè (2sg) kə (3sg.NF) pə (3sg.F) nə (1nsg) ə (2nsg) tə (3nsg) (190)	U	The clitic has the following distribution: "All verbs show agreement for their subject by overt pronominal clitic" (195). "The first exception to obligatory proclitisation is found when a clause consists of a dual pronominal subject in an monovalent clause with no adjunct nominal. In this environment proclitic agreement may be dropped, though it appears that this does not apply equally to all dual subjects." (205) "It seems that this apparent exception to the presence of clitics on verbs in fact conirms their obligatoriness: the only circumstances in which the clitic may be dropped are when a more semantically specified pronoun, with an identical last syllable to the clitic, immediately precedes it; in this case, a purely phonological reduction of two otherwise identical adjacent syllables, attested as a phonological rule elsewhere in the language, occurs (207). (...) The next instance in which a verbal clause may appear without any proclitic agreement is when the subject is both inanimate and there is not a strong degree of affect implied by the verb. (207) "A final instance in which proclitic agreement is not found involves lexicalisations and definitions in which verbs are part of what functions as a nominal compound, but is structurally a syntactic phrase (208). Position variable with adjunct nominal and verb (195). These distributions cannot be interpreted as conditioning contexts, since allomorphs do not exist. Animacy features of participants is not considered GCI.	no	Proclitics never bear stress (82)	no	Clitic not obligatory in the following contexts: "The first exception to obligatory proclitisation is found when a clause consists of a dual pronominal subject in an monovalent clause with no adjunct nominal. In this environment proclitic agreement may be dropped, thugh it appears that this does not apply equally to all dual subjects." (205) "It seems that this apparent exception to the presence of clitics on verbs in fact conirms their obligatoriness: the only circumstances in which the clitic may be dropped are when a more semantically specified pronoun, with an identical last syllable to the clitic, immediately precedes it; in this case, a purely phonological reduction of two otherwise identical adjacent syllables, attested as a phonological rule elsewhere in the language, occurs (207). (...) The next instance in which a verbal clause may appear without any proclitic agreement is when the subject is both inanimate and there is not a strong degree of affect implied by the verb. (207) "A final instance in which proclitic agreement is not found involves lexicalisations and definitions in which verbs are part of what functions as a nominal compound, but is structurally a syntactic phrase (208).

SKOU (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
0.1	LEX/P/N/G	Root (onset), Subject	k/n/h/w/l/Ø (1sg) p/b/m (2sg) k/w/l(3sg.NF) p/t/w/r (3sg.F) t/k/n/w/r (1pl) k/h/w/l/O/ (2pl) t/k/w/r (3pl) (214)	L	Allomorph depending on the verb root (sometimes it is the first consonant of the verb) (214). "One third of all verbs do not inflect by prefixal agreement, and [...] here is no overt prefix for second person nonsingular in all cases, and first person singular in most cases." (241) "lexical stipulativeness involved" (216)	(yes)	Onset of stressed syllable. Can occur as initial element of verbs when P-1 morphs are not present.	(yes)	A root can also be without a lexical onset or agreement prefix, but in few cases only (216).
0.2	LEX/A/T/N/G	Root (vowel), intransitive Subject/Object, Tense, Aspect (228)	Low pitch (past/perfective) Different tonal and vowel alternation patterns for number and gender and person for some verbs (228, 232)	L	Vowel alternation affects only some verbs (228)	yes	First syllable of the root is always stressed (82)	yes	All roots have vowels.
1	A	Intentional, irrealis	Reduplication of the last syllable of the stem (267) (Irrealis, Intentional) (264)	U	No conditioning found except phonological (265).	no	"Under reduplication stress remains with the original root, and not with the reduplicant" (82)	no	
2.1	V	Applicative	na (applicative) (400)	X		no	"With a suffix such as the applicative na we similarly find not change in the position of the stress" (82)	no	

SKOU (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
2.2	PO	Negation (265) (clitic)	ka (negation) (264)	U	No conditioning found.	yes	An independent word, therefore it bears stress.	no	
3	A	Continuous (auxiliary) (265)	i (be.1sg/3sg/2du/2du.F/2pl) me (be.2sg) e (be.3sg.F/3pl) ne (be.1pl) (270f.) (continuous)	G	Inflects for person, gender and number (270).	yes	An independent word, therefore it bears stress.	no	
4	A	Continuous/Intentional (auxiliary) (265)	li (do.1sg/3sg/2du/2du.F/2pl) pi (do.2sg) tue (do.3sg.F) ti (do.1pl/3pl) (270f.) (continuous intentional)	G	Inflects for person, gender and number (270).	yes	An independent word, therefore it bears stress.	no	
5	ADV	Perfective (auxiliary) (269)	loeng (finish) (269)	X	Does not seem to inflect for person, gender and number (269).	yes	An independent word, therefore it bears stress.	no	

SUMERIAN (LCI sample); Jagersma (2010)

Prominence: stress
Sumerian allegedly had a strong stress pattern, deduced from vowel reduction (63). "Sumerian words were stressed on the final syllable" (66) "Forms with clitics were apparently also tressed on the final syllable. If a clitic is attached to the right of a word, the accent shifts to the clitic. The evidence for this comes from forms showing vowel loss." (66)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-11	M/PO	Optative, negative	ha (assertations, wishes, commands) (558) nu (NEG) (J 551, 743)	U	"The use of ha is completely predicta-ble" (558)	no	Never in final posi-tion of the word, therefore not stressed.	no	
-10	A/T/M/PO	Preformatives	ʔi (meaning opaque, used in perfective forms) (548) ʔa (meaning opaque, oc-curs in imperative, stative, and imper-fective forms) (548) Ø (allomorph of ʔi and ʔa) ʔu (relative past, 'when, after' + perfective) (518) na(n) (negative com-mand) (565) ga (imperative first person (hortative)) (569) bara (categorical neg-ative) (557) ši (opaque function, 'non-negative', emphatic assertions) (577-579) na (non-negative, ephatic asser-tions 579) (J 743)	G	ʔi and ʔa are opaque, but they have a distribution according to aspect (548). na(n) is re-stricted to imperfec-tive forms or perfec-tive with a stative meaning (565)	no	Never in final posi-tion of the word, therefore not stressed.	no	
-9	LEX/ADV/V	Sequential	nga (also, then) (J 513, 743)	X		no	Never in final posi-tion of the word, therefore not stressed.	no	
-8	ADV	Ventive	mu/ma (J 498, 743)	X		no	Never in final posi-tion of the word, therefore not stressed.	no	
-7	V	Middle voice	ba (middle voice, passive) (J 487, 494, 743)	X		no	Never in final posi-tion of the word, therefore not stressed.	no	

SUMERIAN (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-6	P/N/G	Indirect object (without or together with the adverbial prefixes P-5 – P-2) (382), Oblique objects (415)	Indirect: ʔ (1sg) e (2sg) r (2sg. indirect; together with P-5 ra) (405ff.) n (3sg.HUM) 9) b (3sg.NHUM) nn (3sg. indirect; together with P-5 nna) (399) mê (1pl) (enê) (2pl) nnê (3pl) (J 382, 743) Oblique: mu (1sg) ri (2sg) nni (3sg.HUM) bi (3sg.NHUM) (mê) (1pl) enê (2pl) nnê (3pl.HUM) (415)	U	No conditioning found except phonological (381ff.; 399ff.)	no	Never in final position of the word, therefore not stressed.	no	
-5	P/G	Indirect object marker	a (to, fo) ra (to, for) (J 381,743)	X	Only if P-1 does not express indirect object (416). Phonological fusion with the following adverbial prefixes.	no	Never in final position of the word, therefore not stressed.	no	
-4	ADV	Comitative	da(with) (J 381 743)	X		no	Never in final position of the word, therefore not stressed.	no	
-3	ADV	Directional	ta (from) ši (to, towards) (J 381, 743)	X		no	Never in final position of the word, therefore not stressed.	no	
-2	ADV	Locative	ni (in, into) e (on, onto) (J 381, 743)	X		no	Never in final position of the word, therefore not stressed.	no	
-1	P/N/G	Transitive subject, direct object, oblique object ('Final person prefixes', 327)	ʔ (1) e (2) n (3.HUM) b 3.NHUM (J 328, 743)	G	"In transitive forms, of the perfective, the final perso-prefix always expresses the transitive subject" (327) In intransitive forms of the perfective, the final person-prefix always expresses the oblique object (327) In transitive forms of the imperfective, the final person-prefix is used to express the direct object (327) in transitive imperfective forms the human prefix can be express the oblique objet if the non-human direct object is left unexpressed. (328)	no	Never in final position of the word, therefore not stressed.	no	The prefixes are not obligatory in intransitive clauses (360).

SUMERIAN (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
0	LEX/ADV/A	Root, pluractionality, imperfective, perfective	Default stem (Perfective/Imperfective) (311) Suppletion/Alternation (Pluractionality) (315).	L	"Some verbs have in the imperfective a special stem which differs in form from the stem in the perfective. Which verb has such a stem and which not is completely unpredictable." (310) The majority of the verbs "have the same stem in the imperfective as in the perfective" (311)	pos.	Can be in final position of the word, therefore potentially stressed	yes	Always present, consists of one or two syllables. (J 309)
1.1	LEX/ADV/A	Pluractionality, imperfective	Partial reduplication (Imperfective forms of some verbs) (312) Full reduplication (Pluractionality) (319) en (Plural marker, lexicalized) (323)	L	Reduplication for imperfective is only found in certain stems (310)	pos.	Can be in final position of the word, therefore potentially stressed	no	
1.2	A	Imperfective	(e)d (Imperfective) (J 370, 743)	U	Regular, except two verbs (du 'go'; bala 'cross') lack this suffix (371)	pos.	Can be in final position of the word, therefore potentially stressed	no	
2	P/N	Subject/ object	en (1/2sg) Ø/e (3sg) enden (1pl) enzen (2pl) eš/enê (3pl.HUM) (J 743)	G	"In intransitive verbal forms, they always refer to the subject" (343) "In transitive forms of the prefective inflection, a person suffix expresses the direct object." (343) "e and enê express the subject in transitive forms of the imperfective inflection, the suffixes Ø and eš are only found in forms of the perfective inflection and in intransitive forms of the imperfective inflection." (343)	pos.	Can be in final position of the word, therefore potentially stressed	(yes)	Zero morpheme in third person. A person suffix is always present; exept in imperative and ga-modal (P-8) (343)
3	ADV	Nominalizer	ʔa (Nominalizer) (591)	X		yes	Is always in final position of the word, therefore always stressed.	no	

SUPYIRE (LCI sample); Carlson (1994)

Prominence: stress
"In the majority of lexical roots (including all verbs) the initial syllable is stressed" (7)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-4	PO	Negation, placed in front of the perfect or progressive auxiliary (379).	ɲyɛ (negation) (379)	G	Only in perfect and progressive (380)	no	"Affixes, clitics, and most other grammatical morphemes (e.g. pronouns, tense-aspect auxiliaries) do not have stress" (7)	no	Only restricted to certain tenses (380)
-3	A/T/M/PO	Negation, Different tenses, aspects, moods. Occupy a distinctive positon in the clause, between the subject and the direct object (307)	Positive: na (progressive) màha (habitual) màha (formal past) ná (remote past) nî (recent past) sí (future) cáá (future) bú/bá (remote (future)) ?a (perfect) sáhá (still, not yet) sí (narrative/sequential asì (habitual/sequential) kù (potential) ta (imperfective imperative) sí (subjunctive) a (imperfective subjunctive) kà (prohibitive) ká (conditional) (308); negative: ɲyɛ à (perfect) ɲyɛ na (progressive) sì (future) càà (future) nà (remote past) nì (recent past) sàhá (still, yet) nàhá (be here) wá (be there) no marking for negationg: màha (habitual) kú (potential) mpyi (past); cannot be negative: narrative/sequential, conditional.	G	Auxiliary conditioned by polarity.	no	"Affixes, clitics, and most other grammatical morphemes (e.g. pronouns, tense-aspect auxiliaries) do not have stress" (7)	yes	"The great majority of cluases have at least one auxiliary, and many combinations of two, three, and even four or more auxiliaries are possible." (307).
-2	NP	Object occurring between auxiliary and verb (307)		X		yes	Has a lexical root as it constitutes its own word.	no	

SUPYIRE (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-1	V/T	Intransitive, future	Toneless nasal stop (intransitive, non-future, perfect, recent past) low-weak mid nasal stop (future)	G	Intransitive Prefix: required by most tense-aspect auxiliaries when immediately preceding the verb. "Only the future auxiliaries (which require the future prefix) and the perfect and recent past (which take no prefix) are not accompanied by this prefix when they occur in intransitive clauses. The itransitive prefix, which consists simply of a toneless nasal, does not actually mark semantic intransitivity, in that it must be used on transitive verbs also, whenever for some reason they are not immediately preceded by their direct object." (127) Future prefix elides when there is a direct object present in which case the L tone of the prefix docks onto the direct object (129)	no	Not the initial root syllable (although syllabic).	no	Intransitive: Not required in transitives and some TMA's (127)
0	LEX/A	Perfective, imperfective	Different tones and vowels determining perfective/imperfective roots (e.g. koo 'cough' perfective kòòlì 'cough', p. 131) lengthening of the vowel (e.g. sú 'pound' perfective, súú 'pound' imperfective, p. 133) vowel alternation (e.g. jya 'break' perfective jyìì 'break' imperfective, p. 133) Consonat mutation (e.g. pa 'come' perfective ma 'come' imperfective, p. 140)	L	Tone/Vowel alternation is lexically conditioned (Paradigms on pp. 130ff). Specific aspects required with specific TMA's (308)	yes	Contains the initial root syllable.	yes	consonant mutation, umlaut, tonal change
1	LEX/ADV/V/A	Imperfective, causative, iterative/intensive, last lexical vowel	lì (imperfective) (130) ni (imperfective) (134) gè (134) Ø (Imperfective) (e.g. p. 135 in variant alternation with ni, otherwise when there is only root(first syllable)-interal alternation like in p. 133) re (Imperfective) (136) different last vowel of the stem (e.g. tuugo 'accompany', perfective tuuge 'accompany' imperfective, p. 132, bubo 'not be well shut' imperfective, bùbi 'not be well shut' perfective, p. 133). gV (Causative in some verbs) (142) lV (non-productive iterative/intensive/participant plurality) (145)	L	75 % of the verbs use li as suffix, it is only found with CVCV roots (130). Different other lexically-conditioned suffixes (130ff.)	no	Not the initial root syllable.	no	Can be also zero (See paradigms 130 ff.)

TUNISIAN ZUWARA BERBER (LCI sample); Mitchell (2009)

Prominence: stress
"Zuaran Berber words and phrases are predominantly paroxytones, i.e. with their penultimate syllable accented." (xi) However, the paradigms of verbs show accent shifts depending on conjugation and category (see morphs where accent is marked).

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-9	PO	Negation	u/wə/Ø (negative) (96-103)	U	No conditioning found except phonological (100-103)	no	This syllable is never stressed.	no	
-8	A/T	Aorist	á (aorist, conjugation 6, 8, 9, 10a, 12b, 13e) a (aorist, conjugation 7) (20) (a) (aorist, conjugation 1, 2, 3, 4a,c, 5, 10b, 11, 12a, 13a-d) (15-16, 21)	L	Whether this morph is stressed depends on the lexical conjugation (27-31)	pos.	Stress possible in conjugations 6, 8, 9, 10a, 12b, 13e. (27-31)	no	
-7	P/N/G	Indirect object	y(ə´) (1sg) k(ə´) (2sg.M) m(ə´) (2sg.F) s(ə´) (3sg) ɣə´n (1pl) kə´n (2pl) sə´n (1pl) (123)	G	Precede the verb in the free-standing aorist, negative past, negative imperative and negative nonpast (114) (131)	pos.	Usually stressed in the aorist positive, not in the aorist past (e.g. examples on p. 126). Can appear word-initially.	no	
-6	P/N/G	Direct object	y (1sg) tt (3g) (1sg) ən(ə´)t (3pl) k (2sg.M) m (2sg.F) ɣə´n (1pl) kə´n (2pl) sə´n (1pl) (123)	G	Precede the verb in the free-standing aorist and the negative past (114). These morphemes do not occur in negative imperative and negative nonpast (131)	pos.	Usually stressed in the aorist positive, not in the aorist past (e.g. examples on p. 126)	no	
-5	P/N/G	Subject	Ø (1sg/3pl) t (2/3sg.F) n (1pl) y (3sg.M) (40)	U	No conditioning found except phonological, with some small irregularities (40-49)	(pos.)	Monoconsonantal; can be part of the stressed syllable that occurs word-initially (e.g. present paradigm on p. 40)	no	Subject marking occurs in every tense, but there are many gaps in this paradigm (1sg, 3pl.m, 3pl.f) that are filled by suffixes in P2. (40).
-4	V	Passive	twa-a/u (66-83) (passive) Can be combined with causative: y-ttwa-s-ə´-h.fəd. 3sg.M-passive-causative-teach.past 'he has been taught' (74)	X		pos.	twa can be stressed (twá), e.g. in conjugation 1 verbs (67)	no	
-3	V	Causative	(ə)s (causative) (52-64)	X		(pos.)	Can be onset of stressed stem like sə´tʕəb (52)	no	

TUNISIAN ZUWARA BERBER (LCI sample); Mitchell (2009)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-2	T	Present, Past, Aorist, Imperative	(ə)t(t)(ə/ə´/V) (present, conjugation 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13); Ø (present, conjugation 1, 6) ə (past conjugation 1, 4, 5, 6, 7, 10b) Ø (past, aorist, imperative, conjugation 2, 3, 8) ə´ (aorist, imperative, conjugation 1, 5, 7, 10b, 11) Ø/ə (aorist, imperative, conjugation 4, 6) ə´(imperative, conjugation 9, 10a, 11) Ø (past, aorist, conjugation 9, 10a) u (past, conjugation 12) á (present, aorist, imperative, conjugation 12a) a (present, aorist, imperative, conjugation 12b) i/u (past, conjugation 13) í/ú/á/i (present, aorist, imperative, conjugation 13) (27-31) ttwa (passive) (64-66)	L	Allomophs depend on the lexical conjugation classes (27-31)	pos.	Stressed in some conjugations (27-31)	no	There are zero allomorphs of every TMA in some conjugations (27-31)
-1	LEX/V	Thematic, reciprocal	t (conjugation 4 marker, can assimilate to following consonant). This marker follows consonant 1 in conjugation 5 (14) (13) m (reciprocal, pejorative) (85-89)	X		(pos.)	Consonantal, can be part of the stressed syllable that occurs word-initially	no	
0.1	LEX	Root, Present	C (first consonant, lexical) gemination of first root consonant (present, conjugation 6b) (29)	L	The gemination only happens with verbs from the conjugation 6b. (29)	(pos.)	Consonantal, can be part of the stressed syllable that occurs word-initially	yes	All verbs have at least one consonant (e.g. ig 'do, make') (31)
0.2	LEX/A/T/M	Present	ə´(present conjugation 1, 6, 7) Ø (non-present conjugaton 1, 6) V´ (all TMA's, all conjugations) (8) ə (non-past, conjugation 7, 12) (20) V´ (past+(present), conjugation 9b, 10, 12, 13) (21) V ((present +) aorist + imperative, conjugation 9b, 10, 12, 13) (21) Ø (all TMA's, conjugation 11)	L	Allomorphs depend on the lexical conjugation classes (27-31)	pos.	See specific allomorphs. (27-31)	no	Segment missing in conjugation 1 and 11
0.3	LEX/V/T	Lexical root consonant(s)	C (second consonant of the root) includes additional consonant of "quadriliteral" words of conjugation 2 (7). Gemination for present in conjugation 1 (27)	L	Past tense inflection: occurs in the second conjugation. Causative derivation (+ s causative prefix) occurs in first conjugation. (5, 52)	(pos.)	Can be coda of stressed syllable at the end of the word, e.g. yugúr (27).	(yes)	Every stem has two consonants, except some mono-consonantal words like yu-f-á 3m-find-past 'he found'or i-g imperative-do 'do' (31)

TUNISIAN ZUWARA BERBER (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
0.4	LEX/A/T/M	Past, present, aorist, imperative, negation	V´1 (accentuated vowel) (past conjugation 1, 5, 6 11), otherwise V1 or Ø; not accentuated, all tenses, all other conjugations (8, 14) ə > a (present, conjugation 3d) Ø (non-past, conjugation 6, 10, 11) V´1> V2 (present, conjugation 3d, 11) Ø (all TMA's, conjugation 12b) í/V´(negative) (90) changes vowel when 1sg and 2pl are suffixed a>i (42)	L	Allomorphs depend on the lexical conjugation classes (27-31; 92-95)	pos.	These vowels can be stressed. See specific allomorphs. (27-31)	no	Segment not present in some conjugations
0.5	LEX	Root	C(C), can elide if an object follows (115)	X		(pos.)	Can be coda of stressed syllable at the end of the word, e.g. yəndíf (31).	no	The third consonant is not present in some words, e.g. yə-šká (complain), however, in this case, the last vowel is lengthened (7)
1.1	LEX/T	Present	a/u/i (present, conjugation 10 and 11) (61-62)	L	Existence of allomorphs depend on the lexical conjugation classes (27-31), allomorph shape itself is conditioned by preceding stem vowel.	no	Does not seem to bear stress, as vowel after the second or third consonant of the root (see conjugation 11 forms on p. 30)	no	
1.2	PO	Negation	i (negative) (91)	L	Allomorphs depend on the conjugation and TMA (91-92)	yes	"Negative suffixation attracts the accent and all negative forms are oxytones." (91)	no	
2	P/N/G	Subject	(ə)d (2sg) ɣ (1sg) (ə)n (3pl.m) n(ə)t (3pl.f) (ə)m (2pl.M) m(ə)t (2pl.F/2pl.F.imperative) Ø (3sg, 1pl) (ə)t (2imperative plural.masculine)	U	No conditioning found except phonological (40-49)	(pos.)	The syllabic morphemes are never stressed although the monoconsonantal can be the coda of a stressed syllable at the end of a word (e.g. mhíɣ, past paradigms on p. 42)	no	Subject marking occurs in every tense, but there are many gaps in this paradigm (3s.m, 3s.f. 1pl) that are filled by suffixes in P-5. (40).

TUNISIAN ZUWARA BERBER (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
3	P/N/G	Indirect object	Stressed or unstressed: (iy)yid/ayid (1sg) (ay)ak (2sg.M) (ay)am (2sg.F) (ay)as (3sg.M.indir.Obj) (ay)anəɣ (1pl) (ay)awən (2pl.M) (ay)akmət (2pl.F) (ay)asən (3pl.M) (ay)asnət (3pl.F) (116) (allomorphs dependent on the final segment of the root) (116)	G	Occurs in the positive past and the positive imperative (114). These morphemes do not occur in negative imperative and negative nonpast (131)	pos.	Depending on the phonologically conditioned allomorphs, stressed or unstressed (116ff.)	no	
4	P/N/G	Direct object	Stressed or unstressed (see paradigms on p. 118-119): (iy)yid/ayid (1sg) (ay)ak (2sg.M) (ay)am (2sg.F) (t)ti(d)/ti (3sg.M) (t)tət/it (3sg.F) (ay)anəɣ (1pl) (ay)awən (2pl.M) (ay)akmət (2pl.F) (t)tən/in (3pl.M) (t)tənt/inət (3pl.F) (116) (allomorphs dependent on the final segment of the root)	G	Occurs in the positive past and the positive imperative, negative imperative and negative nonpast (114, 131)	pos.	Depending on the phonologically conditioned allomorphs, stressed or unstressed (116ff.)	no	
5	PO	Negation	š/Ø (negative) (90)	L	Occurs with certain verbs and not with others, however, these verbs do not form a class (103-110)	(yes)	This suffix seems to be always the coda of a stressed syllable at the end of the word (103ff.)	no	

ULWA (LCI sample); Green (1999)

Prominence: stress
Stress falls on the rightmost syllable of the root (canonically) (58).

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
1	P/N	Subject/ object	yâ (1st) mâ (2nd) yak (1st incl) (113, 114)	G	Only yak is grammatically conditioned; it appears in intransitive classes and marks 1st inclusive subject (105). In transitive clauses, yak refer to objects, in ditransitive to recipients (114). Not found in imperative and infinitives (105).	no	First and second person are long (therefore supposedly stressed; Green does however not mention how stress behaves in inflectional object pronouns), but yak (the grammatically conditioned 1st inclusive morpheme in question) is short, and therefore does not receive stress (114).	no	No 3rd person objects. Also, these pronouns do not appear in intransitive verbs.
0	LEX	Root		X		pos.	Stress falls on the rightmost syllable of the root (canonically) (58).	yes	

ULWA (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
1	A/T/MP/N	Thematic syllable (consonant = vowel or vowel, this syllable is not obligatory), dependent on the specific verb, tense, mood, aspect, person, number (paradigms 119-123).	da, wa, ra, ra, pa, Ø. (119-123).	L	See conjugation classes where verbs are assigned without any grammatical or phonological motivation (104-107). Thematic consonants after the stem. (Garboden 2009: 487) shows that different thematic suffixes are assigned for same/similar semantics across transitivity (497). Some thematic syllables are present in the infinitive and all tenses, whereas ta- doesn't appear for some fused features (e.g. in 3rd plural present) (Green 1999: 122). Pa-thematic syllable can change to wa- in 1st inclusive and vanish in 3rd plural (123).	pos.	Stressed only in vowel-themed roots like yaw-á-rang (go-INFL.TH-3sg.IRR) (59). Other thematic root morphemes are not stressed (58).	(yes)	The Ø-themed verbs are in the minority (106), therefore this position can be considered obligatory.
2	A/T/M/PO/P/N	Present, past, future, perfective, irrealis, negative, obvious, potential, imperative, auditive, intentional, attemptative, agentive (119 ff.)	Various fusional morphemes, e.g. niki (1sg.INF) yang (1sg.PRES) saman (2sg.NEG) (119) person: 1, 2, 3,; number: sg, pl, clusivity: inc, exc.	G	See conjugation classes (119-133). Fusional morphemes depending on tense, mood, aspect, person, number and negation. Non-fusional morphemes: ana (infinitive), i (proximate), ukuh (intentive) î (auditive).	pos.	Imperative and auditive suffix add (additional) stress (59). Inflectional affixes remain unstressed unless there is an underlying long vowel (60), example: á-waní:nah (enter-IMP.NEG.2pl) 'Enter' (60); Here, the verb receives a double accent. This position attracts the stress when there is not a thematic consonant (58), example at-íng (be,say-PERF.1sg) 'I have said/been' (60).	yes	Inflectional affixes are obligatory in this position. See paradigms pp. 90-101. Minimal syllabic requirement one syllable nucleus (like proximative -i in yaw-i 'go-PROX') (96), can be multisyllabic like tasamanna (negative second plural) (99).

URA (LCI sample); Crowley (1999)

Prominence: stress
"Primary stress in Ura is invariably found on the penultimate syllable. In words of four syllables or more, a secondary stress attaches to the preceding syllable but one." The prefixes are rarely stressed as roots are rarely monosyllabic (5.7%, p. 117)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-6	M/P/N/T	Subject, imperative, recent past	Ø (sg.Imperative) ir/ih (pl.Imperative) (157-158). Recent past prefixes: y(au)/yaw (1sg) ki (2sg) (c)(i) (3sg) (q)ur (1pl.inc) qim(i)r (1pl.exc) qir (2pl) (c)ir/(c)ih (3pl) (158, 159). Distant past prefixes: yaum(i) (1sg) kam(i) (2sg) (y)(i) (2sg) (q)ur (1pl.inc.) qimir (1pl.exc) qir (2pl) (c)ir (3pl) (160); Optative prefixes: yaup(i) (1sg) kap(i) (2sg) p(i) (3sg) qisp(i)r (1pl.inc) ? (1pl.exc) qip(i)r (2pl) pir (3pl) (160, 161); Future paradigm: ya(u) (1sg) k(i) (2sg) (c)(u) (2sg) (q)ur(a) (1pl.inc) qimr(a) (1pl.exc) (c)ir(a) (3pl) (161). m/mV/mu/mi (echo suffix marking; when a verb follows another verb in a sentence and the two share the same subject categories) (163)	G	Several interdependent Subject-T/M/A-Suffixes (See paradigms pp. 157-164)	pos.	Stress only if the root is monosyllabic and no other suffixes follow; ergo when this prefix is in penultimate position, e.g. ír-va (pl.IMP-go) 'you (all) go!' (158). Otherwise the epenthetic vowel vanishes and these prefixes can form the onset of the first syllable, p-éni (3sg.OPT-eat) (161)	(yes)	Zero forms in third person (See paradimgs pp. 57-164)
-5	T	Prior past	ehm(i) (prior past) (co-occurs with distant-past-paradigms in P-6) (165)	U	No conditioning found except phonological (164)	no	No evidence for this morpheme being in the penultimate syllabic position, ergo stressed (164)	no	
-4	PO	Negative	etw/et(u) (negative) (165)	U	No conditioning found except phonological (165)	no	No evidence for this morpheme being in the penultimate syllabic position, ergo stressed (165)	no	

URA (LCI sample); Crowley (1999)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-3	ADV	Iterative	oum(i) (iterative) (164)	X	No conditioning found except phonological (164)	no	No evidence for this morpheme being in the penultimate syllabic position, ergo stressed (164-165)	no	
-2	A	Aspectual marker	em(i)/am(i) (no definable meaning) (166), occurs with some categories in P-6 to construct: dependent past, past habitual, present, past continuous. This prefix could mark some kind of imperfective aspect, see Table 6.22. on p. 167.	G	Grammatically conditioned by tense and aspect of the discontinuous construction (166-169). Its presence and form is additionally phonologically conditioned (169)	no	No evidence for this morpheme being in the penultimate syllabic position, ergo stressed (166-171). Monosyllabic roots seem to insert an a- before the root and em-, but monosyllabic roots also use to have object prefixes, see y-ema-da-i (3sg.distantpast-em-hit-3sg) (168)	no	
-1	ADV	Derivative prefixes, reduplication	ovli/amli (simultanitive) omroki/amroki (immediate) esri/asri (random) avyu/amyu (desiderative) (179-180) Full reduplication (183)	X		no	No evidence for these morphemes being in the penultimate syllabic position, ergo stressed (179-180)	no	
0	LEX/ADV/A/T/M/P/N	Basic root: imperative, recent past, distant past, optative, dependent past, past continuous, negative, iterative, purposive, instrumental, derivational, reduplication, causative. Modified root: future, subjunctive, present, habitual (148-150)	Alternations of different segments between basic and modified roots, affecting first vowel (oco > aqo) , first consonant (ivek > ibek), and more (see Table 6.3 on p. 154).	L	"The phonological shape of modified roots vis-à-vis basic roots is determined in part by a categorisation of all verbs as belonging to the class of either weak or strong verbs, and partly on the basis of the phonological shape of the root itself (...) There is an element of unpredictability in the subclassification of verbs in Ura into these two groupings." See Table 6.1. on p. 151.	pos.	Stressed only if the root is part of the penultimate syllable, which is mostly the case (see prefix sets paradigms, p. 159, 160, 161, 162, 164, 167, 170, 173, 174, 175)	yes	

URA (LCI sample); Crowley (1999)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
1	P/N/ADV	Object, transitivity	Object suffixes on ditransitive verbs (give, tell): yau (1sg) ka (2sg) hgi (3sg) kis (1pl.inc) kim (1pl.exc) hgimi (2pl) hgil (3pl) (173, 174). Object suffixes for ta 'hit, kill' and a 'spear': as well as reflexive suffixes yau (1sg) qa (2sg) i (3sg) qis (1pl.inc) qim (1pl.exc) ? (2pl) l (3pl) (174, 176). Object suffix for a larger subset of transitive verbs: l (3pl) (174). Verbs that are partly suffixed have -i in the whole paradigm (transitive?) (175) gi (transitivizer/applicaitve) (181-182) ves (Ameliorative) di (Pejorative) mesiba (thither) mesibenim (hither) belek (outwards) yek (upwards) yip (downwards) (182)	L	"The distribution of forms between the (...) subsets of suffixed and unsuffixed verbs is lexically conditioned, and is not based on either the phonological shape of the verb root, or the semantic categories of the verbs." (176)	pos.	Stressed only if they fill up penultimate syllables, as is the case with 2pl forms (hgími) without following suffix (173) or when post-object suffixes follow (178) (they can appear "after the object suffixes") (177)	no	
2	A/ADV	Perfective, continuative, partitive, misdirective	ye (perfective) la(p) (continuative) wi (partitive) gi (misdirective) (178)	U	No conditioning found (178)	no	Never attracts stress as being the last syllable in the word. (178)	no	
3	P/N	Direct/ reflexive Object	yau (1sg) qa (2sg) Ø (3sg) qis (1pl.inc) qim (1pl.exc) gimi (2pl) Ø (3pl) (175, 176)	L	Appears only in verbs that do not have object suffixes in P1. (176)	yes	These elements are stressed as they are defined as independent words, in contrast to P-1 elements which are not always stressed (cf. pp. 174-175).	no	

YELÎ DNYE (LCI sample); Henderson (1995)

Prominence: stress
"Stress is predictable, falling first on two-syllable words, and first and third on four-syllable words." (5) Since all (portmanteau) inflectional morphemes are single words, almost every position is stressed.

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-4	ADV/PO/P/N/M	Aktionsart, definiteness, deictic, negation, subject	ngmê (indefiniteness) mye (commonality) mê (repetition) n:aa (motion) various (deictic incorporation, e.g. kî or wu- on p. 48/49) negation: daa (unmerged) dî (past.immediate/remote.1sg.sbj) di (past.immediate/remote.2sg.sbj) d:uu (punctiliar/habitual 1/3 subject) dê (continuative/remote past 1/3sg subject) (55) wo-...-pî (contrafactual status) (40)	G	Some prefixes change the stem depending on past and present (see 48 for kî) (48). Negation changes according to aspect, tense and indexation	yes	"Stress is predictable, falling first on two-syllable words, and first and third on four-syllable words." (5) Since all (portmanteau) inflectional morphemes are single words, almost every position is stressed.	no	Subject is only repeated in this position when negation occurs (fusional negative prefix) (55), as such, this is not an obligatory subject position.
-3	A/T/M/P/N	Present/immediate future, near past, remotepast, continuous, punctiliar, indicative, habitual, imperative, first, second or third person subject, number, (35) motion (45). Continuous indicative immediate future and continuous proximal habitual forms are used to indicate motion (46). Negation also part of the portmanteau, see example 125 on page 56. (Clitic)	Used with punctiliar events: future proximal: anî (1sg) anyi (1du) anmî (1pl) anyi (2sg) adpî (2du) anmyi (2pl) Ø (3); immediate future: non-existent; immediate past: dî (1sg) dnye (1du) dpî (3pl) chi (2sg) dpî (2du) dmye (2pl) dê (3); near past/remote past: nî (1sg) nyi (1du) nmî (1pl) nyi (2sg) dpî (2du) nmyi (2pl) Ø (3); Habitual dpî (1sg) dmye (1du) dpî (1sg) dpyi (2sg) dpî (2du) dmye (2pl) dpî (3); Imperative immediate: Ø (thorughout); Imperative deferred: paa (1du/pl) dpî (2/3). (35). See p. 36 for allomorphs used with continuous envents :uu (motion) (45)P. 46: Continuous indicative immediate future and continuous proximal habitual forms are used to indicate motion.	G	Interdependency of categories, especially the main distinction between punctiliar and continuous events (see p. 35 and 36).	pos.	Only stressed if prenucleus is unprefixed.	no	There are still a fair amount of allomorphs that are zero (third person), also non-existent forms like present and immediate future in punctiliar events. (35)
-2	ADV	Motion (44, 45)	naa (motion suffix to prenucleus) (44) mi motion, when the unmarked predicate prenucleus person marking is zero) (45)	X		no	This is a syllabic suffix to a word, and since stress is on the first syllable, it is never stressed.	no	

YELÎ DNYE (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-1	ADV	Incorporation (26)	nouns (non-specific, results in intransitive and continuous predicate) (26)	X		yes	The noun incorporated is an independent word. (see examples on p. 27)	no	
0	LEX/A/T/M	Continuous, punctiliar, 'followed', remote past, punctiliar imperative (33)	Different root alternations for continuous, punctiliar, 'followed', remote past, punctiliar imperative. (33)	L	The occurrence as well as the shape of alternation is determined by the lexicon. (General tendency: Continuous is usually reduplication) (29-34). Formation of roots "regular, almost regular, "paranoid verbs: A class of punctiliar verbs can be termed 'paranoid', because they are sensitive to what follows them in the predicate. One root is used if the postnucleus is zero (...) but a different verb root (termed the 'followed root') is used where the postnucleus is non-zero)" (29).	yes	This is a phonologically independent word (14) and has therefore stress.	yes	This part of the verbal complex is always present: "The predicate nucleus consists of the verb root and sometimes an incorporated noun" (26)
1	ADV/A/T/M/N/P	Subject/Object, Aspect, Tense, Mood	Only third person subject/object exemplified: Intransitives: té (immediate future/present/immediate past continuous) Ø (future/near past/remote past continuous) dmi (future/present/near past punctiliar) dniye (remote past punctiliar) (23) Transitives: ngmê (imediate future/present/immediate past continuous; future immediate past near past punctiliar) ngópu (remote past punctiliar) (24) Object and focus marking (p. 39) See also p. 37 fand 38 for all forms .	G	Since these morphemes are fusional, they are multidependent on the category, see tree of multidependency on pp. 37 and 38.	yes	This is a phonologically independent word (14) and has therefore stress.	no	Several allomorphs are zero (compare pp. 37 and 38)

YUROK (LCI sample); Robins ("R") (1958); Blevins ("B") (2003, 2004)

Pos.	Catego-ries	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-2	A/T/M/PO	Tense, aspect, mood, nega-tion (clitic)	About 48 verbal particles expressing tense, aspect and mood and negation, etc: ho (past) (R 97) kic (perfect) (R 98) wo (past, used after negative parti-cle) (R 98) 'ap (past inceptive) (R 98) 'u (past after kitkwo) (R 98) 'eme/me/ma (past; going and doing) (R 99) ela (past; particular place) (R 99) meo (past; just in time before) (R 99) ki (future/deontic, subor-dinate) (R 99) ku (future; going and doing, subordi-nate) (R 100) kiti (future) (R 100) kitu (going-to-fu-ture) (R 100) kito (intention, desire) (R 101) kesi (future time with anaphoric reference to an event or time already mentioned or implied) (R 101) 'esi (past time, anaphoric) (R 101) mesi (similar to kesi and 'esi but with no time reference) (R 101) 'ocka (present time continuing action) 'ockic (time just past) (R101) wo'ni (Present time and continuing ac-tion) (R 101)	U	No conditioning mentioned: "The particles may be used with any verb form, inflected or noninflected" (R 97)	pos.	Only one par-ticle has a long syllable pa:s (general nega-tive) (R 111)	no	"Yurok verbs may be used in sen-tences, as hasbeen seen, without any preverbal parti-cles, and the syn-tactic functions of the particle + verb complexes are queivalent to those of verbs by them-selves." (R 97)
-1	P/N	Subject	(')ne (1) k'e (2) (')we (3) (R 51)	G	Use of prefix subjects: in some statements or question sen-tences; subordinate verb or verbal group; adverbs, verbal nouns (53), with some pre-verbal particles (R 54)	no	None of the al-lomorphs has a long vowel.	no	Not in all verbal costructions (R 53)
0.1	LEX	Root (initial part of the stem)	numerous lexical elements (B 2004: 332-333)	X		pos.	Can be long (See verbs in B 2004: 332-333)	(yes)	Only few verbs do not have the first formative (See verbs in B 2004: 332-333)
0.2	LEX/ADV	Root, adver-bial (medial part of the stem	o'r /elhk/un/on ('be, of sticklike object)/om/Ø (dif-ferent lexical elements) (B 2004: 332-334)	X		no	No examples where the pre-ceding vowel is long (B 2004: 332-333)	no	Formative can be zero (B 2004: 332-333)

YUROK (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
0.3	LEX/ADV/V/M	Root, pluractionality, transitive, reflexive, imperative, (final part of the stem)	ep (reflexive)/ in (causative)/ (e)t (transitive, goal focus) Ø (e-class formatives) ep' (multiple events) (aa-class formatives) ow/ ('be like') (o-class formatives) k (transitive, event focus)/ Ø/ ech (motion) (oo-class formatives) (B 2004: 332-334) Imperative formation last consonant replaced by a glottal or glottalized (R 45). Glottalization is phonologically predictable from stem formatives (p.46) , and there are few irregular imperatives but the ones that use a different stem alternation than glottalization are few (3 are listed on p. 46), and nasals are not glottalized.	U	Insertion of imperative glottal or trucation phonologically conditioned (R 44-46)	no	No examples where the preceding vowel is long (R 44-47) (B 2004: 332-333)	yes	Formative can be zero; In terms of the imperative, it does not change the number of syllables (R 44-47).
1.1	V	Passive	ey/oy/el (passive) (R 47, 49)	X		no	These morphemes do not have a long vowel. (R 47-49)	no	Passive formation, imperative formation 2 (R 46)
1.2	P/N	Object	p (1sg) c (2sg) s (3sg) Ø (1pl; here, the passive morpheme in P1 is solely used) c' (2pl when 1sg is subject; otherwise zero or 2sg. passive form is used) s' (3pl only when 1sg is subject, otherwise Ø) (R 70)	G	Object forms are dependent on subject person/number (R 70)	(no)	Only consonantal onset. (R 70) Cannot occur at the end of the word since suffixes must follow.	no	Only in transitive verbs. "Specific bipoersonal forms only fill certain places in the system, namely, those in most frequent use. The other places are filled either by forms derived from the passive stem but used in syntactic structures appropriate to an active verb, or by forms of the unipersonal conjugation." (R 69)

YUROK (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
2.1	P/N/M	Subject	O -Class: Cok' (1sg) Co'm/Co:'m (2sg) C'/'C/Co'm/cok'w/Co'l/Co' (3sg) Coh (1pl) Co'w (2sg) Coł (3pl) (34) os (imperative singular) (R 44) ekw (imperative plural) (R 46)	L	Choice between the classes can be predicted by some lexical formatives (B 2004: 332-333) but it's not always consistent, since there are also zero formatives in all classes. (B 2004: 336) Shape of allomorphs in o-class lexically conditioned (R 33).	pos.	In monosyllabic o-class verbs, the vowel of all (except 3sg) affixes is lengthened (and therefore stressed) (R 34).	(yes)	Obligatory, however, third person singular can have only glottalization as overt realization which is zero-syllabic (R 33)
2.2	M/P/N	Subject	E-Class: Cek' (1sg) Ce'm (2sg) C'/'C (3sg) Coh (1pl) Cu' (2pl) Ceł (3pl) (34) es (imperative singular) (44)	L	Choice between the classes can be predicted by some lexical formatives (B 2004: 332-333) but it's not always consistent, since there are also zero formatives in all classes. (B 2004: 336)	no	These morphemes do not have a long vowel.	(yes)	Obligatory, however, third person singular can have only glottalization as overt realization which is zero-syllabic (R 33)

ZACATEPEC CHATINO (LCI sample); Villard (2015)

Prominence: tonal variability
No stress pattern recorded. Instead, there is a 6-tone contrast: a/à (low-slightly falling) ā (mid, leveled) á (high, leveled) ǎ (rise from low to mid-high) a̋ (rise from low to super-high) (145)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-3	M/A	Potential, habitual, progressive, completive	k/tyi/Ø+laminalization of root consonant (potential) nti (habitual) ntV(k) (progressive) nk(w)V(y) (completive) (287)	G	Allomorphs are conditioned by either valency or phonology of the first consonant of the stem (see p. 288 for a summary)	pos.	Only have low or middle tone (See 287 and paradigms pp. 290-314)	yes	Aspect and mood prefixes are obligatory (276)
-2	LEX/V	Causative	u (Thematic causative depending on verb stem) (289) /ix (296) (Thematic causative depending on verb stem)	X		pos.	Always low or middle tone, fuses with prefixes in P-3 (292, 295, 297, 301, 309).	no	
-1	LEX/V	Transitivity	t/s (Transitive. The alternation seems thematic. E.g. 'raise it' uses s-causative, whereas 'close it' has t-causative (334) (Intransitive) (276)	X		(no)	Coda of P-3 or P-2 ergo low or middle. (321, 324, 326, 328, 330, 333). Never appears at the end or beginning of a word.	no	
0	LEX/A/M/P/N	Root, completive, progressive, habitual, potential, subject	Tone alternation for tense, mood, aspect and person. The tone alternations are classified into lexically conditioned classes. 10 There are 10 Aspectual classes (317-350) interacting with 9 Classes for person (365-401). The second and first person tone is predictable from the default tonal class (3rd person) (365). This makes the person alternation lexically predictable, because the default form (3rd person) is also lexically predictable. person: 1,2,3; number: sg, pl, clusivity: inc, exc.	L	Tone classes are lexical (177ff.) E.g. Class 4: "The aspectual tonal pattern of this conjugation class cannot be predicted on the basis of the completive tonal pattern alone. Which of the two patterns a verb shows seems to be unpredictable" (326)	yes	Several tone patterns, implementing all tonal contrasts (See paradigms 324-350)	yes	Always present (276)
1	P/N	Subject	nan (1pl.inc) wa (1pl.exc) wan (2pl) neʔ (3pl) (416)	U	No conditioning found except phonological (416)	no	Unmarked for tone (416)	no	Only present in plural (416)

ǂHÒÃ (UCI sample); Collins & Gruber (2014)

Prominence: tonal variability
Tone variability 1 tone = no; 2-4 tone = possible 5 tone = yes

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-7	PO	Negation	ǀ’hö̈õ (negation) (23)	U	No conditioning found (23-24).	no	Negation has only one tone (23)	no	
-6	M	Potential	ǀqe (can) (47f.)	U	No conditioning found (47).	no	can' has only one tone (47)	no	
-5	T	Past, future	i (recent past) (25) ǁke (distant past) (27) ǀkù (hesternal past) (26) ča (Immediate Future) (33) qo (ordinary future, events after today) (33) o (general future) (33) kx’ǒ (remote future) (36)	U	Recent past: No conditioning found (25). Hesternal past: No conditioning found (26). Distant past: no conditioning found (27). Past tense markers are not obligatory in the past when there are subordinate clauses (28) or adverbials (27) Futures: No conditioning found (32ff.). Futures follow past tense markers, see past future: ma i ča č'eo ya 1sg PAST FUT do 3sg 'I was going to do it.' (26) but precede progressive marking (36).	pos.	Mid tone: ' i (recent past) (25) ǁke (distant past) (27) ča (Immediate Future) (33) qo (ordinary future, events after today) (33) o (general future) (33) ; Low tone: ǀkù (hesternal past) (26); Rising (low-mid) tone: kx’ǒ (remote future) (36)	no	
-4	ADV/A	Habitual, direction	ǀna (habitual, 'sit') (25) 'kyà (go, motion toward an intended action) (45f.) Occurs right before the progressive marker (45-46). ǀkama (dubitive) (46) (can also occur before the negation marker in P-7	U	Habitual: No conditioning found (24-25), occurs with imperfective.	pos.	Mid tone: 'ǀna (habitual, 'sit') (25) ǀkama (dubitive) (46); Low tone: 'kyà (go, motion toward an intended action) (45f.)	no	
-3	ADV/A	Progressive, itinerate, positional orverbs	a (progressive/habitual≈imperfective) (23, 24) present time interpretation for unbounded states (38) nà (itinerate) (43-44), positional preverbs, e.g. !gò'a (stand) (43) ǂgà’a (recline) (44) (can’t be used with imperfective or progressive)	U	Imperfective: No conditioning found (23,24)	pos.	Mid tone: 'a (progressive/habitual≈imperfective) (23, 24) present time interpretation for unbounded states (38); Low tone: nà (itinerate) (43-44), positional preverbs, e.g. !gò'a (stand) (43) ǂgà’a (recline) (44)	no	

ǂHÒÃ (cont.)

Pos.	Categories	Functions	Morphs/Features	Cond.	Explanation Cond.	Prom.	Explanation Prom.	Obl.	Explanation Obl.
-2	ADV	Clause type	ma/m̀ (relative imperfective) (107) ka (subordination) (181) qù (subjunctive) (185)	X		pos.	Mid tone: ma (relative imperfective) (107) ka (subordination) (181); Low tone: m̀ (relative imperfective) qù (subjunctive) (185)	no	
-1	ADV	Pluractionality	kí (event plurality) (65) kí (causative) (165) kì (passive) (164) kìí (passive causative) (165)	X		pos.	High tone: kí (event plurality) (65) kí (causative) (165); Low tone: kì (passive) (164); Low/high tone: kìí (passive causative) (165)	no	
0	LEX	Lexical root	Can show suppletion for plural (58ff.)	X		yes	Roots show all tonal constrasts.	yes	
1.1	ADV	Pluractionality, serial verb	ǁqò (spatial distributive) (63) ču (repetitive) (65) n|ne (spatial distributivity) (67), reduplication (sequential repetition) (68-69). Several serial verbs |'o (put.in) (167) describing directionals č (come) kyú (rise) ǂ'o (exit) khoa (arrive) kà khy'a (bring) (169) !xào (remove) |'úšo (enter) (170), resultative lexemes (171), completed lexemes(171), simultaneous lexemes (171), sequential lexemes (172) benefactive šú (173)	X		yes	Since this is a lexical position where further verbs can occur, there is full tonal variability.	no	
1.2	A	Perfective	a (perfective) (38-39)	U	No conditioning found. However, states that are constructed as results of an action have a present time interpretation ('hunger killed me' > I am hungry 'I have seen Titi > I know Titi' 'I have become tired > I am tired) (39)	no	a (perfective) has mid-tone (38).	no	

References

Abbott, C. (2000). Oneida. Languages of the world: Materials 301. München: Lincom Europa.

Abboub, N., Nazzi, T., & Gervain, J. (2016). Prosodic grouping at birth. Brain and Language, 162, 46-59.

Ackerman, F., & Malouf, R. (2013). Morphological organization: The low conditional entropy conjecture. Language, 89(3), 429-464.

Aikhenvald, A. Y., & Dixon, R. M. (Eds.). (2007). Grammars in contact: A cross-linguistic typology (Vol. 4). OUP Oxford.

Aksu-koç, A. A., & Slobin, D. I. (1985). The Acquisition of Turkish. In D. I. Slobin (Ed.), The Crosslinguistic Study of Language Acquisition: Vol. 1. The Data. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Albin, D. D., & Echols, C. H. (1996). Stressed and word-final syllables in infant-directed speech. Infant Behavior and Development, 19(4), 401-418.

Allen, G. D., & Hawkins, S. (1978). The Development of Phonological Rhythm. In A. Bell & J. B. Hooper (Eds.), Syllables and Segments (pp. 173-85). Amsterdam: North Holland Publishing Company.

Allen, M., Badecker, W., & Osterhout, L. (2003). Morphological analysis in sentence processing: An ERP study. Language and Cognitive Processes, 18(4), 405-430.

Altman, G., & Carter, D. M. (1989). Lexical stress and lexical discriminability: Stressed syllables are more informative, but why? Computer Speech & Language, 3, 265-275.

Anderson, S. R. (1982). Where's Morphology? Linguistic Inquiry, 13, 571-612.

Anderson, S. R. (1992). A-morphous Morphology. Cambridge: Cambridge University Press.

Anderson, S. R. (2015). Dimensions of morphological complexity. In Baerman, M., Michelson, K. & Corbett G. G. (Eds.), Understanding and measuring morphological complexity (pp. 1-26). Oxford University Press.

Andrason, A. (2014). Language complexity–an insight from complex system theory. International Journal of Language and Linguistics, 2(2), 74-89.

Andrews, S. (1989). Frequency and neighborhood size effects on lexical access: activation or search? Journal of Experimental Psychology, 15, 802-814.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: lexical similarity or orthographic redundancy? Journal of Experimental Psychology, 18, 234-254.

Anttila, R. (1972). An introduction to historical and comparative linguistics. New York: Macmillan.

Arkadiev, P., & Gardani, F. (Eds.). (2020). The Complexities of Morphology. Oxford University Press.

Aronoff, M. (1998). Isomorphism and monotonicity: Or the disease model of morphology. In Lapointe, S., Brentari, D., & Farrell, P. (Eds.), Morphology and its Relation to Phonology and Syntax (pp. 411-418). Stanford, CA: CSLI Publications.

Aronoff, M. (2015). Thoughts on morphology and cultural evolution. In Bauer, L., Körtvélyessy, L., & Štekauer, P. (Eds.), Semantics of Complex Words (pp. 277-288). Cham: Springer. doi:10.1007/978-3-319-14102-2_13

Ashby, W. J. (1977). Clitic inflection in French: An historical perspective. Amsterdam: Rodopi.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. Journal of Memory and Language, 37(1), 94-117.

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., & Blevins, J. (2019). The discriminative lexicon: a unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. Complexity, 2019:4895891. doi:10.1155/2019/4895891

Baerman, M., Brown, D., & Corbett, G. G. (Eds.). (2015). Understanding and measuring morphological complexity. Oxford University Press, USA.

Bakker, D. (2011). Language sampling. In The Oxford handbook of linguistic typology.

Balling, L. W., & Baayen, R. H. (2012). Probability and surprisal in auditory comprehension of morphologically complex words. Cognition, 125(1), 80-106.

Barnes, J. (2006). Strength and weakness at the interface: Positional neutralisation in phonetics and phonology. Berlin: Walter de Gruyter.

Batchelor, R. E., & Chebli-Saadi, M. (2011). A reference grammar of French. Cambridge University Press.

Baumann, S., & Cangemi, F. (2020). Integrating phonetics and phonology in the study of linguistic prominence. Journal of Phonetics, 81, 100993.

Beckman, J. N. (1998). Positional faithfulness (Doctoral dissertation). University of Massachusetts, Amherst.

Beckman, M. E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In Keating, P. (Ed.), Papers in Laboratory Phonology (pp. 7-33). Cambridge: Cambridge University Press.

Beckner, C., Blythe, R., Bybee, J. L., Christiansen, H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., & Schoenemann, T. (2009). Language Is a Complex Adaptive System: Position Paper. Language Learning, 59, 1-26.

Bell, A. (1978). Language samples. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), Universals of Human Language, Vol. 1: Method and Theory (pp. 123-156). Stanford: Stanford University Press.

Berent, I., & Shimron, J. (1997). The representation of Hebrew words: evidence from the contour principle. Cognition, 64, 39-72.

Berent, I., Vaknin, V., & Marcus, G. F. (2007). Roots, stems, and the universality of lexical representations: Evidence from Hebrew. Cognition, 104(2), 254-286.

Berger, H. (1998). Die Burushaski-Sprache von Hunza und Nager (Vol. 13). Teil I: Grammatik. Wiesbaden: Otto Harrassowitz Verlag.

Berko, J. (1958). The child's learning of English morphology. Word, 14(2-3), 150-177.

Berman, R. A. (1977). Natural Phonological Processes at the One-Word Stage. Lingua, 43, 1-21.

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. Journal of Experimental Psychology: General, 117(1), 21.

Bickel, B., & Nichols, J. (2007). Inflectional morphology. In T. Shopen (Ed.), Language Typology and Syntactic Description (pp. 169-240). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511618437.003

Bickel, B., & Nichols, J. (2013a). Fusion of Selected Inflectional Formatives. In M. S. Dryer & M. Haspelmath (Eds.), The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/20, Accessed on 2021-12-02.)

Bickel, B., & Nichols, J. (2013b). Inflectional Synthesis of the Verb. In M. S. Dryer & M. Haspelmath (Eds.), WALS Online (v2020.3) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7385533 (Available online at http://wals.info/chapter/22, Accessed on 2023-08-30.)

Bisang, W. (2014). On the strength of morphological paradigms. In M. Miestamo, K. Sinnemäki, & F. Karlsson (Eds.), Paradigm change: In the Transeurasian languages and beyond (pp. 23-60). Amsterdam: John Benjamins.

Blake, B. J. (1969). The Kalkatungu Language: A Brief Description. Canberra: Australian Institute for Aboriginal Studies.

Blevins, J. (2003). Yurok syllable weight. International Journal of American Linguistics, 69(1), 4-24.

Blevins, J. (2004). Evolutionary phonology: The emergence of sound patterns. Cambridge: Cambridge University Press.

Blevins, J. (2005). Yurok verb classes. International Journal of American Linguistics, 71(3), 327-349.

Blevins, J. (2009). Structure-preserving sound change: A look at unstressed vowel syncope in Austronesian. In A. Adelaar & A. Pawley (Eds.), Austronesian historical linguistics and culture history: A festschrift for Robert Blust (pp. 33-49). Canberra: Pacific Linguistics.

Blevins, J. P. (2007). Conjugation classes in Estonian. Linguistica Uralica, 43(4), 250-267.

Bloomfield, L. (1933). Language. New York: Holt, Rinehart and Winston.

Blythe, R. A., & Croft, W. (2012). S-Curves and the mechanisms of propagation in language change. Language, 88(2), 269-304. http://www.jstor.org/stable/23251832

Booij, G. E. (2005). The grammar of words: An introduction to linguistic morphology. Oxford: Oxford University Press.

Booij, G. (2006). Inflection and Derivation. In K. Brown (Ed.), Encyclopedia of Language & Linguistics (2nd ed., Vol. 5, pp. 654-661). Oxford: Elsevier.

Bornkessel-Schlesewsky, I., Choudhary, K. K., Witzlack-Makarevich, A., Bickel, B., Malchukov, A., & Richards, M. (2008). Bridging the gap between processing preferences and typological distributions: Initial evidence from the online comprehension of control constructions in Hindi. Linguistische Arbeitsberichte, (86), 397-436.

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S.-Y., Pascual, L., Pêcheux, M.-G., Ruel, J., Venuti, P., & Vyt, A. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean and American English. Child Development, 75(4), 1115-1139.

Bouchon, C., Floccia, C., Fux, T., Adda-Decker, M., & Nazzi, T. (2015). Call me Alix, not Elix: Vowels are more important than consonants in own-name recognition at 5 months. Developmental Science, 18(4), 587-598.

Bowern, C. (2012). A grammar of Bardi. De Gruyter Mouton.

Bowern, C. L. (2004). Bardi verb morphology in historical perspective. (Doctoral dissertation). Harvard University.

Boye, K., & Harder, P. (2012). A usage-based theory of grammatical status and grammaticalization. Language, 1-44.

Boyland, J. T. (1996). Morphosyntactic Change in Progress: A Psycholinguistic Approach (Doctoral Dissertation). University of California.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. Phonetica, 49(3-4), 155-180.

Brown, P. (1998). Children's first verbs in Tzeltal: evidence for an early verb category. Linguistics, 36(4), 713-754. https://doi.org/10.1515/ling.1998.36.4.713

Brown, P., Pfeiler, B., de León, L., Pye, C., Bavin, E., & Stoll, S. (2013). The acquisition of agreement in four Mayan languages. In E. Bavin & S. Stoll (Eds.), The acquisition of ergativity (pp. 271-306).

Bucci, J., Perrier, P., Gerber, S., & Schwartz, J. L. (2019). Vowel reduction in Coratino (South Italy): Phonological and phonetic perspectives. Phonetica, 76(4), 287-324.

Bülow, L., de Bot, K., & Hilton, N. (2017). Zum Nutzen der Complex Dynamic Systems theory (CDST) für die Erforschung von Sprachvariation und Sprachwandel [Conference presentation]. Presented at the IGDD, Stuttgart.

Büring, D. (2010). Towards a typology of focus realization. In M. Zimmermann & C. Féry (Eds.), Information Structure (pp. 177-205). Oxford: Oxford University Press.

Burquest, D. A., & Payne, D. L. (1993). Phonological analysis: A functional approach. Dallas, TX: Summer Institute of Linguistics.

Butler, B. A. (2014). Deconstructing the Southeast Asian sesquisyllable: A gestural account. (Doctoral dissertation). Cornell University.

Bybee, J. L. (1985). Morphology: A Study of the Relation Between Meaning and Form (Vol. 9). John Benjamins Publishing.

Bybee, J. L. (1995). Regular morphology and the lexicon. Language and Cognitive Processes, 10(5), 425-455.

Bybee, J. L. (2001). Phonology and Language Use. Cambridge University Press.

Bybee, J. L. (2006). From Usage to Grammar: The Mind's Response to Repetition. Language, 82(4), 711-733.

Bybee, J. L. (2010). Language, Usage and Cognition. Cambridge: Cambridge University Press.

Bybee, J. L. (2015). Language change. Cambridge University Press.

Bybee, J. L., & Hopper, P. (Eds.) (2001). Frequency and the Emergence of Linguistic Structure. John Benjamins Publishing.

Bybee, J. L., & Thompson, S. (1997). Three frequency effects in syntax. In Annual Meeting of the Berkeley Linguistics Society, Vol. 23(1), 378-388.

Bybee, J. L., & Moder, C. L. (1983). Morphological classes as natural categories. Language, 251-270.

Bybee, J. L., & Newman, J. E. (1995). Are stem changes as natural as affixes? Linguistics, 33(4), 633-654.

Bybee, J. L., & Slobin, D. I. (1982). Rules and schemas in the development and use of the English past tense. Language, 58(2), 265-289.

Bybee, J. L., Chakraborti, P., Jung, D., & Scheibman, J. (1998). Prosody and segmental effect some paths of evolution for word stress. Studies in Language, 22(2), 267-314.

Bybee, J. L., Perkins, R. D., & Pagliuca, W. (1994). The evolution of grammar: Tense, aspect, and modality in the languages of the world (Vol. 196). Chicago: University of Chicago Press.

Campbell, L., Kaufman, T., & Smith-Stark, C. (1986). Meso-America as a linguistic area. Language, 62(3), 530-70.

Carlson, R. (1994). A grammar of Supyire (Vol. 14). Berlin/New York: Walter de Gruyter.

Carstairs-McCarthy, A. (2010). The evolution of morphology (Vol. 14). Oxford: Oxford University Press.

Chao, Y. (1968). A grammar of spoken Chinese. Berkeley and Los Angeles: University of California Press.

Chee, M. R. (2017). A longitudinal cross-sectional study on the acquisition of Navajo verbs in children aged 4 years 7 months through 11 years 2 months (Doctoral dissertation) University of New Mexico.

Cheek, A., Cormier, K., Repp, A., & Meier, R. P. (2001). Prelinguistic gesture predicts mastery and error in the production of early signs. Language, 77(2), 292-323.

Chen, M. Y. (2004). Tone sandhi: Patterns across Chinese dialects (Vol. 92). Cambridge University Press.

Chersoni, E., Blache, P., & Lenci, A. (2016, December). Towards a distributional model of semantic complexity. In COLING Workshop on Computational Linguistics for Linguistic Complexity (pp. 12-22).

Chomsky, N. (1980). Rules and representations. New York: Pantheon Books.

Chomsky, N. (1986). Knowledge of language. Its nature, origin and use. New York: Praeger.

Chomsky, N. (2007). Biolinguistic explorations: Design, development, evolution. International Journal of Philosophical Studies, 15(1), 1-21.

Clancy, P. M. (1985). The acquisition of Japanese. In D. A. Slobin (Ed.), The crosslinguistic study of language acquisition (Vol. 1: The data) (pp. 373-524). New Jersey: Lawrence Erlbaum.

Clark, E. V. (2017). Morphology in language acquisition. The handbook of morphology, 374-389.

Cole, J., Hualde, J. I., Smith, C. L., Eager, C., Mahrt, T., & de Souza, R. N. (2019). Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. Journal of Phonetics, 75, 113-147.

Collins, C., & Gruber, J. S. (2014). A Grammar of ǂHõã. Köln: Rüdiger Köppe.

Comrie, B. (1980). Morphology and word order reconstruction: Problems and prospects. In J. Fisiak (Ed.), Historical morphology (pp. 83-96). Den Haag: Mouton.

Comrie, B. (1988). Coreference and conjunction reduction in grammar and discourse. In J. Hawkins (Ed.), Explaining language universals (pp. 186-210). Oxford: Blackwell.

Corbett, G. (2000). Number. Cambridge: Cambridge University Press.

Cotterell, R., Mielke, S. J., Eisner, J., & Roark, B. (2018). Are all languages equally hard to language-model? arXiv preprint arXiv:1806.03743.

Courtney, E. H., & Saville-Troike, M. (2002). Learning to construct verbs in Navajo and Quechua. Journal of child language, 29(3), 623-653.

Cristofaro, S. (2019). Taking diachronic evidence seriously: Result-oriented vs. source-oriented explanations of typological universals. In K. Schmidtke-Bode, N. Levshina, S. M. Michaelis, & I. A. Seržant (Eds.), Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence (pp. 25-46). Berlin: Language Science Press.

Croft, W. (1994). Semantic universals in classifier systems. Word, 45(2), 145-171.

Croft, W. (1999). Adaptation, optimality and diachrony. Zeitschrift für Sprachwissenschaft, 18(2), 206-208.

Croft, W. (2000). Explaining language change: an evolutionary approach. Harlow: Longman.

Croft, W. (2001). Radical Construction Grammar: Syntactic Theory in Typological Perspective. Oxford and New York: Oxford University Press.

Croft, W. (2003). Typology and universals. Cambridge University Press.

Croft, W. (2009). Toward a Social Cognitive Linguistics. In V. Evans & S. Pourcel (Eds.), New Directions in Cognitive Linguistics (pp. 395-420). Amsterdam and Philadelphia: John Benjamin.

Croft, W. (2013). An evolutionary model of language change and language structure. Available online: https://www.unm.edu/~wcroft/Papers/ELC2-Chap02.pdf

Crowley, T. (1999). Ura: A disappearing language of southern Vanuatu. Dept. of Linguistics, Research School of Pacific Studies, The Australian National University.

Cysouw, M., & Dediu, D. (2013). Some structural aspects of language are more stable than others: A comparison of seven methods. PloS one, 8(1), e55009.

Dahl, Ö. (2004). The growth and maintenance of linguistic complexity (Studies in Language Companion Series, Vol 71). Amsterdam/Philadelphia: John Benjamins Publishing.

Dahl, Ö. (2008). An exercise in a posteriori language sampling. Language Typology and Universals, 61(3), 208-220. https://doi.org/10.1524/stuf.2008.0021

De Boysson-Bardies, B., & Vihman, M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. Language, 67(2), 297-319.

De Busser, R. (2015). The Influence of Social, Cultural, and Natural Factors on Language Structure: An Overview. In R. De Busser & R. J. LaPolla (Eds.), Language Structure and Environment: Social, Cultural, and Natural Factors (pp. not specified). Amsterdam: John Benjamins.

De Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyper-articulation. The journal of the acoustical society of America, 97(1), 491-504.

De Jong, K., Beckman, M. E., & Edwards, J. (1993). The interplay between prosodic structure and coarticulation. Language and Speech, 36, 1972-2212.

De Jonge, B., & Nieuwnhuijsen, D. (2009). Formación del paradigma pronominal de las formas de tratamiento. In C. Company Company (Ed.), Sintaxis histórica de la lengua española: Segunda parte: La frase nominal, vol. 2 (pp. 1595–1671). México D. F.: Fondo de Cultura Económica.

De Lacy, P. (1999). Tone and prominence. Rutgers Optimality Archive, 333.

De Reuse, W. (2005). Template-Induced Prefix Deletion, Accommodation, and Shunting in Western Apache. Alaska Native Language Center Research Papers, 5, 95-109.

DeCapua, A. (2017). Grammar for teachers. New York: Springer.

Denk, L. (2019). Ordering towards disorder: Explaining the stability of non-layered morpheme structure in Athabascan languages. Studies in Language, 43(4), 800-849.

Denk, L. (2020). Argument marking and phonological alignment: an alternative explanation for the retention of morphological complexity in Ket. SN Social Sciences, 1(1), 39.

DiCristo, A. (1999). Vers une modélisation de l'accentuation du français (première partie). J. French Lang. Stud. 9, 143-179.

Dixon, R. M. W. (1994). Ergativity. Cambridge: Cambridge University Press. DOI:10.1017/CBO9780511611896

Dixon, R. M.W. (1997). The rise and fall of languages. Cambridge University Press.

Dixon, R. M.W. (2001). The Australian linguistic area. Areal diffusion and genetic inheritance: problems in comparative linguistics, 64-104.

Dixon, R. M. W. (2010). Basic linguistic theory volume 1: Methodology (Vol. 1). Oxford University Press.

Donohue, M. (2004). A grammar of the Skou language of New Guinea. Unpublished draft, downloaded from: https://pure.mpg.de/rest/items/item_402710/component/file_402709/content

Dressler, W. U. (1989). Prototypical Differences between Inflection and Derivation. Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung, 42, 3-10.

Dressler, W. U. (2005a). Morphological Typology and First Language Acquisition: Some Mutual Challenges. In G. Booij, E. Guevara, A. Ralli, S. Sgroi, & S. Scalise (Eds.), Morphology and Linguistic Typology, On-line Proceedings of the Fourth Mediterranean Morphology Meeting (MMM4) Catania 21-23 September 2003, University of Bologna.

Dressler, W. U. (2005b). Word-formation in natural morphology. In P. Hohenhaus, P. Štekauer, & R. Lieber (Eds.), Handbook of word-formation, 267-284.

Dryer, M. S. (2011a). Feature 26 A: Prefixing vs. suffixing in inflectional morphology. The World Atlas of Language Structures Online. Munich: Max Planck Digital Library. https://wals.info/chapter/26

Dryer, M. S. (2011b). The evidence for word order correlations. Linguistic Typology, 15, 335-380.

Dryer, M. S. (1989). Large linguistic areas and language sampling. Studies in Language, 13, 257-292.

Dryer, M. S., & Haspelmath, M. (Eds.) (2013). The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info (Accessed on 2021-12-01.)

Easterday, S. (2017). Highly complex syllable structure: A typological study of its phonological characteristics and diachronic development (Doctoral Dissertation). University of New Mexico.

Echols, C. H., & Newport, E. L. (1992). The role of stress and position in determining first words. Language acquisition, 2(3), 189-220.

Eddington, D., & Lachler, J. (2006). A computational analysis of Navajo verb stems. Empirical and Experimental Methods in Cognitive/Functional Research. CSLI Publications.

Ellis, R. (1999). Item versus system learning: Explaining free variation. Applied linguistics, 20(4), 460-480.

Ellis, R. L., & Jones, M. R. (2009). The role of accent salience and joint accent structure in meter perception. Journal of Experimental Psychology: Human Perception and Performance, 35(1), 264-280.

Everett, D. L. (1986). Pirahã. In D. C. Derbyshire & G. K. Pullum (Eds.), Handbook of Amazonian Languages (1), 200-325. Berlin: Mouton de Gruyter.

Falk, S. (2014). On the notion of salience in spoken discourse-prominence cues shaping discourse structure and comprehension. TIPA. Travaux interdisciplinaires sur la parole et le langage, (30).

Faust, N., & Lampitelli, N. (2016). Allomorphy—its logic and limitations: opening remarks from the guest editors. Morphology, 26(3-4), 229-234.

Fedden, S. (2011). A grammar of Mian. De Gruyter Mouton.

Fenk-Oczlon, G., & Fenk, A. (2008). Complexity trade-offs between the subsystems of language. In Language Complexity: Typology, Contact, Change, 43-65. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Fenk-Oczlon, G., & Fenk, A. (2014). Complexity trade-offs do not prove the equal complexity hypothesis. Poznan Studies in Contemporary Linguistics, 50(2), 145-155.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. Developmental psychology, 27(2), 209-221.

Fernández-Dobao, A., & Herschensohn, J. (2020). Present tense verb morphology of Spanish HL and L2 children in dual immersion: Feature Reassembly revisited. Linguistic Approaches to Bilingualism, 10(6), 775-804.

Fifer, W. P., & Moon, C. M. (1994). The role of mother's voice in the organization of brain function in the newborn. Acta Paediatrica, 83, 86-93.

Flegontov, P., Changmai, P., Zidkova, A., Logacheva, M. D., Altınışık, N. E., Flegontova, O., Gelfand, M. S., Gerasimov, E. S., Khrameeva, E. E., Konovalova, O. P., Neretina, T., Nikolsky, Y. V., Starostin, G., Stepanova, V. V., Travinsky, I. V, Tříska, P., & Tatarinova, T. V. (2016). Genomic study of the Ket: a Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry. Scientific Reports, 6, https://doi.org/10.1038/srep20768

Flemming, E. (2004). Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), Phonetically-based phonology, 232-276.

Fortson, B. W. IV. (2003). An Approach to Semantic Change. In B. D. Joseph & R. D. Janda, The Handbook of Historical Linguistics, 648-666.

Fougeron, C. (1999). Prosodically conditioned articulatory variations: a review. UCLA Working Papers in Phonetics, 97, 1-74.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. Journal of the Acoustical Society of America, 101(6), 3728-3740.

Foulet, L. (1921). Comment ont évolué les formes de l'interrogation. Romania, 47(186/187), 243-348.

Frota, S. (2014). Prosody and focus in European Portuguese: Phonological phrasing and intonation. London and New York: Routledge.

Fruchter, J., Stockall, L., & Marantz, A. (2013). MEG masked priming evidence for form-based decomposition of irregular verbs. Frontiers in Human Neuroscience, 7, 798. 10.3389/fnhum.2013.00798

Frumhoff, P., Echols, C. H., & Newport, E. L. (1992). Perceptual Salience and Operating Principles for Language Acquisition: The Effects of Stress and End of Word, ms., University of Texas, Austin.

Gabelentz, G. v. d. (2016). Die Sprachwissenschaft: Ihre Aufgaben, Methoden und bisherigen Ergebnisse. OCLC 14, 950016913. doi:10.26530/oapen_611696

Garagnani, M., Shtyrov, Y. Y., & Pulvermuller, F. (2009). Effects of attention on what is known and what is not: MEG evidence for functionally discrete memory circuits. Frontiers in human neuroscience, 3, 10.

Geeraerts D., Grondelaers S., Bakema P. (1994). The Structure of Lexical Variation. Meaning, Naming and Context. Mouton de Gruyter, Berlin.

Gennari, S., & Poeppel, D. (2003). Processing correlates of lexical semantic complexity. Cognition, 89(1), B27-B41.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. Center for the Study of Reading Technical Report; no. 257.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity and early word learning. In Bowerman, M., & Levinson, S. (Eds.), Language acquisition and conceptual development (215-256). Cambridge University Press.

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In Gelman, S. A., & Brynes, J. P. (Eds.), Perspectives on thought and language: Interrelations in development (225-277). London: Cambridge University Press.

Gerner, M., & Ling, Z. (2020). Zero morphemes in paradigms. Studies in Language, 44(1), 1-26.

Gerz, D., Vulić, I., Ponti, E. M., Reichart, R., & Korhonen, A. (2018). On the relation between linguistic typology and (limitations of) multilingual language modeling. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (pp. 316-327).

Gierut, J. A. (2007). Phonological Complexity and Language Learnability. American Journal of Speech-Language Pathology, 16(1), 6.

Giraudo, H., & Dal Maso, S. (2016). The salience of complex words and their parts: Which comes first? Frontiers in psychology, 7(1778). doi: 10.3389/fpsyg.2016.01778

Givón, T. (1971). Historical syntax and synchronic morphology: An archeologist's field trip. Chicago Linguistic Society 7, 394-415.

Givón, T. (1979). On understanding grammar. New York: Academic Press.

Givón, T. (2009). The genesis of syntactic complexity. The Genesis of Syntactic Complexity, 1-384.

Givón, T., & Shibatani, M. (Eds.). (2009). Syntactic complexity: Diachrony, acquisition, neuro-cognition, evolution (Vol. 85). John Benjamins Publishing.

Gleitman, L. R., & Wanner, E. (1982). Language Acquisition: The State of the State of the Art. In Wanner, E., & Gleitman, L. R. (Eds.), Language Acquisition: The State of the Art. Cambridge, England; Cambridge University Press.

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. Language Learning and Development, 1(1), 23-64.

Goldin-Meadow, S., Seligman, M. E., & Gelman, R. (1976). Language in the two-year old. Cognition, 4(2), 189-202.

Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the "natural order of L2 morpheme acquisition," in English: A meta-analysis of multiple determinants. Language learning, 51(1), 1-50.

Goldsmith, J. (1987, September). Tone and accent, and getting the two together. In Annual Meeting of the Berkeley Linguistics Society (Vol. 13), 88-104.

Good, J. (2011). The typology of templates. Language and Linguistics Compass, 5(10), 731-747.

Gordon, M. (2011). Stress: Phonotactic and Phonetic Evidence. In The Blackwell Companion to Phonology (Eds M. Oostendorp, C.J. Ewen, E. Hume, & K. Rice). https://doi.org/10.1002/9781444335262.wbctp0039

Gordon, M., & Roettger, T. (2017). Acoustic correlates of word stress: A cross-linguistic survey. Linguistics Vanguard, 3(1).

Gourley, J. W., & Catlin J. (1978). Children's Comprehension of Grammatical Structures in Context. Journal of Psycholinguistic Research, 7(6), 419-434.

Grabe, E., Post, B., & Watson, I. (1999, August). The acquisition of rhythmic patterns in English and French. In Proceedings of the 14th International Congress of Phonetic Sciences (1201-1204). Berkeley, CA: University of California.

Granier-Deferre, C., Bassereau, S., Ribeiro, A., Jacquet, A. Y., & DeCasper, A. J. (2011). A melodic contour repeatedly experienced by human near-term fetuses elicits a profound cardiac reaction one month after birth. PLoS One, 6(2), e17304.

Green, T. M. (1999). A lexicographic study of Ulwa (Doctoral dissertation). Massachusetts Institute of Technology.

Greenberg, J. H. (1966). Language universals, with special reference to feature hierarchies. (Janua Linguarum, Series Minor, 59.) The Hague: Mouton.

Gross, M., Say, T., Kleingers, M., Clahsen, H., & Münte, T. F. (1998). Human brain potentials to violations in morphologically complex Italian words. Neuroscience Letters, 241(2-3), 83-86.

Grünloh, T., Lieven, E., & Tomasello, M. (2011). German Children Use Prosody to Identify Participant Roles in Transitive Sentences. Cognitive Linguistics 22, 393-419.

Gundel, J. K. (1988). Universals of Topic-Comment structure. In Hammond, M., Moravcsik, E. A., & Wirth, J. R. (Eds.), Studies in Syntactic Typology (209-239). Amsterdam: John Benjamins Publishing.

Haiman, J. (1994). Ritualization and the development of language. In Pagliuca, W. Perspectives on Grammaticalization (pp. 3-28). Amsterdam: John Benjamins.

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016, November). Cultural shift or linguistic drift? comparing two computational measures of semantic change. In Proceedings of the conference on empirical methods in natural language processing. Conference on empirical methods in natural language processing.

Hammarström, H. (2010). The status of the least documented language families in the world. Language Documentation and Conservation, 4, 177-212.

Hammarström, H., Forkel, R., Haspelmath, M., Bank, S. (2021). Glottolog 4.4. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://glottolog.org, Accessed on 2021-12-01.)

Hantgan, A. (2013). Aspects of Bangime phonology, morphology, and morphosyntax (Doctoral dissertation). Indiana University.

Hare, M., & Elman, J. L. (1995). Learning and morphological change. Cognition, 56(1), 61–98. https://doi.org/10.1016/0010-0277(94)00655-5

Harley, H. (2006). English Words: A Linguistic Introduction. Malden: Blackwell.

Harms, R. T. (1962). Estonian Grammar. (Indiana University Publications. Uralic and Altaic Series 12.) Bloomington: Indiana University.

Harris, J. (1969). Spanish phonology. Cambridge, Mass: MIT Press.

Harris, J. (2005). Vowel reduction as information loss. In Carr, P., Durand, J., & Ewen, C. J. (Eds.), Headhood, elements, specification and contrastivity: Phonological papers in honour of John Anderson (pp. 119-132). Amsterdam: John Benjamin.

Haspelmath, M. (1999). Optimality and diachronic adaptation. Zeitschrift für Sprachwissenschaft, 18(2), 180-205.

Haspelmath, M. (2008). Frequency vs. iconicity in explaining grammatical asymmetries. Cognitive Linguistics, 19(1), 1-33. doi:10.1515/COG.2008.001

Haspelmath, M. (2010). Comparative concepts and descriptive categories in crosslinguistic studies. Language, 86(3), 663-687.

Haspelmath, M. (2019). Can cross-linguistic regularities be explained by constraints on change? In Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence (1-23). Language Science Press.

Haspelmath, M. (2020). The morph as a minimal linguistic form. Morphology, 30(2), 117-134.

Haspelmath, M.; König, E.; Oesterreicher, W.; et al., eds. (2001). The European linguistic area: Standard Average European. In Language typology and language universals (1492-1510). Berlin: de Gruyter.

Hawkins, J. A. (2004). Efficiency and complexity in grammars. Oxford University Press. Oxford.

Hawkins, J. A. (2007). Processing typology and why psychologists need to know about it. New ideas in Psychology, 25(2), 87-107.

Hawkins, J. A. (2014). Cross-linguistic variation and efficiency. Oxford: Oxford University Press. DOI:10.1093/acprof:oso/9780199664993.001.0001

Hawkins, J. A., & Gilligan, G. (1988). Prefixing and suffixing universals in relation to basic word order. Lingua, 74(2-3), 219-259.

Hay, J. (2001). Lexical frequency in morphology: is everything relative? Linguistics, 39(6), 1041-1070. https://doi.org/10.1515/ling.2001.041

Hayden, R. E. (1950). The relative frequency of phonemes in general American English. Word, 6(3), 217-223. https://doi.org/10.1080/00437956.1950.11659381

Hayes, B. (1995). Metrical Stress Theory: Principles and Case Studies. Chicago: The University of Chicago Press.

Heath, J. (1984). Functional grammar of Nunggubuyu. Canberra: Australian Institute of Aboriginal Studies.

Heath, J., & Hantgan, A. (2018). A grammar of Bangime. De Gruyter Mouton.

Heine, B., & Song, K. A. (2011). On the grammaticalization of personal pronouns1. Journal of Linguistics, 47(3), 587-630.

Heitmeier, M., Chuang, Y. Y., & Baayen, R. H. (2021). Modeling morphology with linear discriminative learning: Considerations and design choices. Frontiers in psychology, 12, 720713.

Heylighen, F. (2016). Stigmergy as a universal coordination mechanism I: Definition and components. Cognitive Systems Research, 38, 4-13.

Hellenthal, A. C. (2010). A grammar of Sheko. Netherlands Graduate School of Linguistics.

Henderson, J. (1995). Phonology and grammar of Yele, Papua New Guinea. Dept. of Linguistics, Research School of Pacific Studies, The Australian National University.

Hendrikse, R., & van Zweel, H. (2010). A phylogenetic and cognitive perspective on linguistic complexity. Southern African Linguistics and Applied Language Studies, 28(4), 409-422.

Henri, F. Stump, Gregory, & Tribout Delphine (2020). Derivation and the morphological complexity of three French-based creoles. In Arkadiev, P., & Gardani, F. (Eds.), The Complexities of Morphology. Oxford University Press.

Herbert, R. K. (1986). Language universals, markedness theory, and natural phonetic processes. Trends in Linguistics, Studies and Monographs 25. Berlin, New York, and Amsterdam: Mouton de Gruyter.

Hildebrandt, K. A. (2003). Manange tone: Scenarios of retention and loss in two communities (Doctoral dissertation). University of California, Santa Barbara.

Himmelmann, N. (2000). A typology of typologies. Sprachtypologie und Universalienforschung, 53, 5-12.

Hochmann, J. R., Benavides-Varela, S., Nespor, M., and Mehler, J. (2011). Consonants and vowels: different roles in early language acquisition. Developmental science, 14(6), 1445-1458.

Hoekstra, E., and Versloot, A. P. (2019). Factors promoting the retention of irregularity. Morphology, 29(1), 31-50.

Höhle, B., Schmitz, M., Santelmann, L. M., & Weissenborn, J. (2006). The recognition of discontinuous verbal dependencies by German 19-month-olds: Evidence for lexical and structural influences on children's early processing capacities. Language Learning and Development, 2(4), 277-300.

Hoijer, H. (1945). Navaho Phonology. University of New Mexico Publications in Anthropology, (No. 1). University of New Mexico Press.

Holton, G. (2000). The Phonology and Morphology of the Tanacross Athabaskan Language (Doctoral Dissertation). University of California, Santa Barbara.

Hull, D. (1988). Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science. Chicago: University of Chicago Press.

Huttenlocher, J. (1974). The origins of language comprehension. In Solso, R. L. (Ed.), Theories of cognitive psychology: The Loyola Symposium. Lawrence Erlbaum.

Hymes, D. H. (1956). Na-Dene and positional analysis of categories. American Anthropologist, 58, 624-638. https://doi.org/10.1525/aa.1956.58.4.02a00040

Ingram, D. (1976). Phonological Analysis of a Child. Glossa, 10, 3-27.

Inkelas, S. (1993). Nimboran position class morphology. Natural Language and Linguistic Theory, 11, 559-624.

Jagersma, B. (2010). A descriptive grammar of Sumerian (Doctoral dissertation). Leiden University.

Jaker A., & Howson, P. (2022). An Acoustic Study of Tetsǫ̨́t’ıné Stress: Iambic stress in a quantity-sensitive tone language. Phonology, 1-39.

Järvikivi, J., Bertram, R., & Niemi, J. (2006). Affixal salience and the processing of derivational morphology: The role of suffix allomorphy. Language and cognitive processes, 21(4), 394-431.

Joanisse, M. F., & McClelland, J. L. (2015). Connectionist perspectives on language learning, representation and processing. Wiley Interdisciplinary Reviews: Cognitive Science, 6(3), 235-247.

Justus, T., Larsen, J., de Mornay Davies, P., & Swick, D. (2008). Interpreting dissociations between regular and irregular past-tense morphology: Evidence from event-related potentials. Cognitive, Affective, & Behavioral Neuroscience, 8(2), 178-194.

Justus, T., Yang, J., Larsen, J., de Mornay Davies, P., & Swick, D. (2009). An event-related potential study of cross-modal morphological and phonological priming. Journal of Neurolinguistics, 22(6), 584-604.

Kako, E. T. (2005). Information sources for noun learning. Cognitive Science, 29, 223-260.

Kapatsinski, V. M. (2005). To scheme or to rule: Evidence against the Dual-Mechanism Model and the Rule-Based Learner. In Annual Meeting of the Berkeley Linguistics Society, 31(1), 193-204.

Kapatsinski, V. M. (2018). Changing minds changing tools: From learning theory to language acquisition to language change. Cambridge, Massachusetts: MIT Press.

Kari, J. (1989). Affix positions and zones in the Athapaskan verb complex: Ahtna and Navajo. International Journal of American Linguistics, 55(4), 424-454.

Karzon, R. G. (1985). Discrimination of Polysyllabic Sequences by One-to-Four-Month-Old Infants. Journal of Experimental Child Psychology, 39, 326-342.

Kathol, A. (2000). Linear syntax. Oxford: Oxford University.

Kavitskaya, D. (2002). Triggers and alternations in compensatory lengthening. 21st West Coast Conference of Formal Linguistics. Somerville: Cascadilla Press.

Keating, P. A., Cho, T., Fougeron, C., & Hsu, C. (2003). Domain-initial strengthening in four languages. In Lo-cal J., Ogden, R., & Temple, R. (Eds.) Papers in laboratory phonology, Vol. 6: Phonetic interpretation (145-163). Cambridge, UK: Cambridge University Press.

Keating, P., Linker, W., & Huffman, M. (1983). Patterns in allophone distribution for voiced and voiceless stops. Journal of phonetics, 11(3), 277-290.

Keller, R. E. (1978). The German language. New Jersey: Humanities Press.

Kenstowicz, M. (1996). Quality-sensitive stress. Rivista di Linguistica, 9(1), 157-187.

Kidder, E. (2008). Tone, intonation, stress and duration in Navajo. Coyote Papers: Working Papers in Linguistics, Linguistic Theory at the University of Arizona (Vol. 16). http://hdl.handle.net/10150/126405

Kielar, A., & Joanisse, M. F. (2010). Graded effects of regularity in language revealed by N400 indices of morphological priming. Journal of cognitive neuroscience, 22(7), 1373-1398.

Kilarski, M. (2014). Complexity in the history of language study. Poznan Studies in Contemporary Linguistics, 50(2), 157-168.

Kimball, G. D. (1991). A descriptive grammar of Koasati (Louisiana). University of Nebraska Press.

Klein, H. E. M. (1973). A grammar of Argentine Toba: Verbal and nominal morphology. Columbia University.

Koehlinger, K., Van Horne, A. O., Oleson, J., McCreery, R., & Moeller, M. P. (2015). The role of sentence position, allomorph, and morpheme type on accurate use of s-related morphemes by children who are hard of hearing. Journal of Speech, Language, and Hearing Research, 58(2), 396-409.

Koenig, J. P., & Michelson, K. (2015). Morphological complexity à la Oneida. In Baerman, M., Michelson, K. & Corbett G. G. (Eds.), Understanding and measuring morphological complexity (69-92). Oxford University Press.

Kohler, K. J. (2008). The perception of prominence patterns. Phonetica, 65, 257-269.

Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. Problems of information transmission, 1-7.

Kõrgvee, K. (2001). Lapse sõnavara areng vanuses 1;8-2;1. [A child's lexical development, aged 1;3-2;1.] (Unpublished undergraduate thesis) Tartu University.

Krauss, M. E. (1965). Proto-Athapaskan-Eyak and the problem of Na-Dene II: morphology. International Journal of American Linguistics, 31(1), 18-28.

Kruspe, N. (2004). A grammar of Semelai. Cambridge University Press.

Kusters, W. (2003). Linguistic Complexity; the Influence of Social Change on Verbal Inflection. PhD Dissertation, University of Leiden.

Kusters, W. (2008). Complexity in linguistic theory, language learning and language change. In Miestamo M., Sinnemäki K. & Karlsson F. (Eds.), Language complexity: Typology, contact, change (pp. 3-22). Amsterdam: Benjamins.

Kuznetsova, N., & Verkhodanova, V. (2019). Phonetic realisation and phonemic categorisation of the final reduced corner vowels in the Finnic languages of Ingria. Phonetica, 76(2-3), 201-233.

Kuznetsova, N., & Anderson, C. (2020). Vowel reduction and loss: challenges and perspectives. Italian Journal of Linguistics, 3-17.

Lacheret-Dujour, A., & Beaugendre, F. (1999). La Prosodie du Français (French Prosody) (CNRS, Paris), 1-353.

Lacrampe, S. (2014). Lelepa: Topics in the grammar of a Vanuatu language (Doctoral Dissertation). Australian National University.

Ladányi, E., Kovács, Á. M., & Gervain, J. (2020). How 15-month-old infants process morphologically complex forms in an agglutinative language?. Infancy, 25(2), 190-204.

Ladefoged, P. (2001). Vowels and consonants. Phonetica, 58(3), 211-212.

Ladefoged, P., & Disner, S. F. (2012). Vowels and consonants. John Wiley & Sons.

Ladefoged, P., & Maddieson, I. (1996). The sounds of the world's languages. Oxford: Blackwell Publishers.

Lakoff, G. (1987). Women, Fire, and Dangerous Things. What Categories Reveal about the Mind. Chicago: University of Chicago Press.

Langacker, R. W. (1991). Foundations of Cognitive Grammar (Vol. 2: Descriptive Application). Stanford: Stanford University Press.

Langacker, R. W. (2008). Cognitive Grammar: A Basic Introduction. Oxford, UK: Oxford University Press.

Langacker, R. W. (2019). Morphology in cognitive grammar. In Audring J & Masini F, eds. The Oxford Handbook of Morphological Theory. Oxford, UK: Oxford University Press, 346-364.

Lavoie, L. (2001). Consonantal strength: Phonological patterns and phonetic manifestations. New York: Garland.

Lecanuet, J. P., & Schaal, B. (1996). Fetal sensory competencies. European Journal of Obstetrics & Gynecology and Reproductive Biology, 68, 1-23.

Lee, S. N. (1996). A grammar of Iranian Azerbaijani (Doctoral dissertation). University of Sussex.

Lehiste, I. (1970). Suprasegmentals. Cambridge, Massachusetts: MIT Press.

Leminen, A., Smolka, E., Dunabeitia, J. A., & Pliatsikas, C. (2019). Morphological processing in the brain: The good (inflection), the bad (derivation) and the ugly (compounding). Cortex, 116, 4-44.

Lieber, R. (2009). Introducing Morphology. New York: Cambridge University Press.

Linares, R. E., Rodriguez-Fornells, A., & Clahsen, H. (2006). Stem allomorphy in the Spanish mental lexicon: Evidence from behavioral and ERP experiments. Brain and Language, 97(1), 110-120.

Liu, L., & Kager, R. (2015). Understanding phonological acquisition through phonetic perception: The influence of exposure and acoustic salience. Phonological Studies, 18, 51-58.

Loevenbruck, H. (1999). An investigation of articulatory correlates of the Accentual Phrase in French. In Ohala, J., Hasegawa, Y., Ohala, M., & Granville, D. (Eds.), Proceedings of the 14th international congress of phonetic sciences, San Francisco, 667-670.

Loevenbruck, H. (2000). Effets articulatoires de l'emphase contrastive sur la Phrase Accentuelle en français. Actes des 23ème Journées d'Etude sur la Parole, Aussois, 165-169.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. PloS one, 5(1), e8559. doi: 10.1371/journal.pone.0008559

MacDonald, M., & Christiansen, M. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). Psychological Review, 109(1), 35-54.

Macnamara, J. (1972). Cognitive basis of language learning in infants. Psychological Review, 79, 1-13.

MacWhinney, B. (Ed.) (1999). The Emergence of Language. Lawrence Erlbaum Associates.

Maddieson, I. (1984). Patterns of Sounds. Cambridge: Cambridge University Press.

Maddieson, I. (2005). Correlating phonological complexity: data and validation. UC Berkeley PhonLab Annual Report, 1(1).

Maddieson, I. (2009). Calculating phonological complexity. Approaches to phonological complexity, 85-110.

Maddieson, I. (2013a). Absence of common consonants. In Dryer, M. S., & Haspelmath, M. (Eds.) The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (http://wals.info/chapter/18, Accessed on 2021-05-18.)

Maddieson, I. (2013b). Tone. In Dryer, M. S., & Haspelmath, M. (Eds.) The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (http://wals.info/chapter/13, Accessed on 2021-05-18.)

Maddieson, I. (2013c). Consonant Inventories. In Dryer, M. S., & Haspelmath, M. (Eds.) The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (http://wals.info/chapter/1, Accessed on 2021-05-18.)

Maddieson, I. (2013d). Vowel Quality Inventories. In Dryer, M. S., & Haspelmath, M. (Eds.) The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (http://wals.info/chapter/2, Accessed on 2021-05-18.)

Maddieson, I. (2013e). Voicing and Gaps in Plosive Systems. In Dryer, M. S., & Haspelmath, M. (Eds.) WALS Online (v2020.3). Zenodo. https://doi.org/10.5281/zenodo.7385533 (Available online at http://wals.info/chapter/5, Accessed on 2023-09-16.)

Maganga, C., & Schadeberg, T. C. (1992). Nyamwezi: Grammar, vocabulary and texts. Köln: Rüdiger Köppe.

Mansfield, J. (2017). Prosodic words in cyclic derivation: The strange case of Murrinhpatha compound verbs. Morphology, 27(3), 359-382.

Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. Cognitive psychology, 29(3), 189-256.

Marinova-Todd, S. H., Marshall, D. B., & Snow, C. E. (2000). Three misconceptions about age and L2 learning. TESOL quarterly, 34(1), 9-34.

Marslen-Wilson, W., & Tyler, L. K. (1998). Rules, representations, and the English past tense. Trends in cognitive sciences, 2(11), 428-435.

Martin, A., & Peperkamp, S. (2020). Phonetically natural rules benefit from a learning bias: a re-examination of vowel harmony and disharmony. Phonology, 37(1), 65-90.

Marzi, C., Ferro, M., Nahli, O., Belik, P., Bompolas, S., & Pirrelli, V. (2018, May). Evaluating inflectional complexity crosslinguistically: a processing perspective. In Calzolari N, Choukri K., Cieri C., Declerck T., Goggi S., Hasida K., Isahara H., Maegaard B., Mariani J., Mazo H., Moreno A., Odijk J., Piperidis S. & Tokunaga T. (Eds.), Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018) (pp. 3860-3866). Miyazaki, Japan.

Matisoff, J. (1990). Bulging monosyllables: Areal tendencies in Southeast Asian diachrony. In Hall, K. (Ed.), Proceedings of the 16th Annual Meeting of the Berkeley Linguistics Society (pp. 543-559).

Matisoff, J. (2003). Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction. University of California Publications in Linguistics.

Maton, K., & Doran, Y. J. (2017). Semantic density: A translation device for revealing complexity of knowledge practices in discourse, part 1—wording. Onomázein, 46-76.

Matras, Y. (2020). Language contact [Second Edition]. Cambridge University Press.

Mayr, E. (1978). Evolution. Scientific American, 239, 46-55.

McClelland, J. & Bybee, J. (2007). Gradience of Gradience: A reply to Jackendoff., 24(4), 437-455. https://doi.org/10.1515/TLR.2007.019

McDonough, J. (2003). The Navajo Sound System. Dordrecht: Kluwer Academic Publishers (now Springer).

McWhorter, J. H. (2001). The world's simplest grammars are creole grammars. Linguistic typology, 5(2-3), 125-166.

Miceli, L., & Dench, A. (2017). The Areal Linguistics of Australia. In The Cambridge Handbook of Areal Linguistics (pp. 732-757). Cambridge University Press.

Mielke, S. J., Cotterell, R., Gorman, K., Roark, B., & Eisner, J. (2019). What kind of language is hard to language-model? arXiv preprint arXiv:1906.04726.

Miestamo, M. (2008). Grammatical complexity in a cross-linguistic perspective. In Miestamo M., Sinnemäki K. & Karlsson F. (Eds.), Language complexity: Typology, contact, change, (Studies in Language Companion Series 94), Amsterdam: Benjamins, 23-41.

Miestamo, M., Bakker, D., & Arppe, A. (2016). Sampling for variety. Linguistic Typology, 20(2), 233-296.

Miestamo, Matti, Kaius Sinnemäki, and Fred Karlsson (2008). Language complexity: Typology, contact, change. Amsterdam: John Benjamins.

Millar, R. M., & Trask, L. (2015). Trask's historical linguistics. Routledge.

Mitchell, T. F. (2009). Zuaran Berber (Libya): Grammar and Texts. Köln: Rüdiger Köppe.

Mithun, M. (1988). Lexical categories and the evolution of number marking. In M. Hammond and M. Noonan (Eds.), Theoretical Morphology: Approaches in Modern Linguistics, 211-234. Tokyo/ Toronto: Academic Press Inc. Harcourt, Brace, Jovanovich.

Mithun, M. (1989). The acquisition of polysynthesis. Journal of Child Language, 16(2), 285-312.

Mithun, M. (2011). "Grammaticalization and explanation," In Narrog, H. and Heine, B. (Eds.), The Oxford handbook of grammaticalization, 177-192. Oxford: Oxford University Press.

Mondloch, J. L. (2013). Basic K'ichee' Grammar: 38 lessons. University Press of Colorado.

Moravcsik, E. A. (2011). Explaining language universals. In Jae Jung Song (Ed.), The Oxford handbook of language typology, 69-89. Oxford University Press.

Moravcsik, E. A. (2012). Introducing language typology. Cambridge University Press.

Morris, J., & Holcomb, P. J. (2005). Event-related potentials to violations of inflectional verb morphology in English. Cognitive Brain Research, 25(3), 963-981.

Morris, J., & Stockall, L. (2012). Early, equivalent ERP masked priming effects for regular and irregular morphology. Brain and language, 123(2), 81-93.

Moscoso del Prado, F. (2011). The mirage of morphological complexity. In Proceedings of the Annual Meeting of the 33rd Cognitive Science Society, 3524-3529.

Mowrey, R. and Pagliuca, W. (1995). The reductive character of articulatory evolution. Rivista di linguistica, 7, 37-124.

Mufwene, S. S. (2008). Language Evolution: Contact, Competition and Change, London and New York: Continuum.

Mufwene, S. S. (2014). Language Ecology, Language Evolution, and the Actuation Question. In T.A. Åforli & B. Mæhlum (Eds.), The Sociolinguistics of Grammar, 130-36. Amsterdam: John Benjamins.

Mufwene, S. S. (2018). Language contact and evolutionary linguistics: An African (ist)'s and creolist's perspective.'. The dynamics of language, 36-51.

Mufwene, S. S., Coupé, C., & Pellegrino, F. (Eds.) (2017). Complexity in language: Developmental and evolutionary perspectives. Cambridge University Press.

Mukherjee, Joybrato (2005). English Ditransitive Verbs. Aspects of theory, description and a usage-based model. Amsterdam / New York: Rodopi.

Münte, T. F., Say, T., Clahsen, H., Schiltz, K., & Kutas, M. (1999). Decomposition of morphologically complex words in English: Evidence from event-related brain potentials. Cognitive Brain Research, 7(3), 241-253.

Nam, Y. J., & Polka, L. (2013, June). The role of acoustic/perceptual salience in directional asymmetry in infant stop/fricative contrast perception. In Proceedings of Meetings on Acoustics (Vol. 19, No. 1). AIP Publishing.

Naranjo, M. G., & Becker, L. (2021). Coding efficiency in nominal inflection: expectedness and type frequency effects. Linguistics Vanguard, 7(s3).

Neef, M. (2000). Morphologische und syntaktische Konditionierung. In Booij G., Lehmann C. & Mugdan J., Kesselheim W. & Skopeteas S. (Eds.), Morphology: An International Handbook on Inflection and Word-Formation (Vol. 1), 473-484.

Neef, M. (2000b). Phonologische Konditionierung. In Booij G., Lehmann C. & Mugdan J., Kesselheim W. & Skopeteas S. (Eds.), Morphology: An International Handbook on Inflection and Word-Formation (Vol. 1), 463-473.

Nespor, M., Peńa, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. Lingue e linguaggio, 2(2), 203-230.

Nevins, A. (2011). Phonologically conditioned allomorph selection. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice (Eds.), The Blackwell companion to phonology (pp. 2357-2382). Oxford: Blackwell.

Newman, A. J., Ullman, M. T., Pancheva, R., Waligura, D. L., & Neville, H. J. (2007). An ERP study of regular and irregular English past tense inflection. NeuroImage, 34(1), 435-445.

Newmeyer, Frederick J. (2007). Linguistic typology requires crosslinguistic formal categories. Linguistic Typology, 11, 133-157.

Nichols, J. (1992). Linguistic diversity in space and time. Chicago, IL: University of Chicago Press.

Nichols, J. (1993). Ergativity and linguistic geography. Australian Journal of Linguistics, 13(1), 39-89.

Nichols, J. (2007). What, if anything, is typology? Linguistic Typology, 11(1).

Nichols, J. (2009). Linguistic complexity: A comprehensive definition and survey. In Sampson G., Gil D. & Trudgill P. (Eds.), Language complexity as an evolving variable (pp. 110-125). Oxford: Oxford University Press.

Nichols, J. (2011a). Causativization and contact in Nakh-Daghestanian. In Proceedings of the 37th Annual Meeting of the Berkeley Linguistics Society, Special Session on Languages of the Caucasus, 68-80.

Nichols, J. (2011b). Ingush grammar. University of California Publications in Linguistics, Vol. 143. Berkeley: University of California Press.

Nichols, J. (2019), Why is gender so complex? Some typological considerations. In F. Di Garbo, B. Olsson & B. Wälchli (Eds.), Grammatical gender and linguistic complexity: Volume I: General issues and specific studies. Studies in Diversity Linguistics, no. 26. Berlin: Language Sciences Press, 63-9.

Nikolaev, A., Lehtonen, M., Higby, E., Hyun, J., & Ashaie, S. (2018). A facilitatory effect of rich stem allomorphy but not inflectional productivity on single-word recognition. Applied Psycholinguistics, 39(6), 1221-1238.

Norcliffe, E., Harris, A. C., & Jaeger, T. F. (2015). Cross-linguistic psycholinguistics and its critical role in theory development: Early beginnings and recent advances. Language, Cognition and Neuroscience, 30(9), 1009-1032.

Nordlinger, R., & Caudal, P. (2012). The tense, aspect and modality system in Murrinh-Patha. Australian Journal of Linguistics, 32(1), 73-113.

O'Connor, M. (1987). Topics in Northern Pomo Grammar. (Doctoral Dissertation). University of California, Berkeley.

Oller, D. K., & Rydland, R. E. (1974). Note on the stress preferences of young English-speaking children. Unpublished manuscript, Mailman Center for Child Development, Miami, FL.

Orsolini, M., Fanari, R., & Bowles, H. (1998). Acquiring regular and irregular inflection in a language with verb classes. Language and cognitive processes, 13(4), 425-464.

Ortega, L. (2015). Syntactic complexity in L2 writing: Progress and expansion. Journal of second language writing, 29, 82-94.

Osgood, C. E., & Hoosain, R. (1974). Salience of the word as a unit in the perception of language. Perception & Psychophysics, 15, 168-192. https://doi.org/10.3758/BF03205845

Overall, S. E. (2017). A Grammar of Aguaruna (Iiniá Chicham) (Vol. 68). De Gruyter Mouton.

Pallotti, G. (2019). An Approach to Assessing the Linguistic Difficulty of Tasks. Journal of the European Second Language Association, 3(1), 58-70. DOI: https://doi.org/10.22599/jesla.61

Park, H. H., Zhang, K. J., Haley, C., Steimel, K., Liu, H., & Schwartz, L. (2021). Morphology Matters: A Multilingual Language Modeling Analysis. Transactions of the Association for Computational Linguistics, 9, 261-276. doi: https://doi.org/10.1162/tacl_a_00365

Partanen, E., Kujala, T., Tervaniemi, M., & Huotilainen, M. (2013). Prenatal music exposure induces long-term neural effects. PloS one, 8(10), e78946.

Pellegrino, F., Marsico, E., Chitoran, I., & Coupé, C. (Eds.). (2009). Approaches to phonological complexity (Vol. 16). Walter de Gruyter.

Penke, M., Weyerts, H., Gross, M., Zander, E., Münte, T. F., & Clahsen, H. (1997). How the brain processes complex words: an event-related potential study of German verb inflections. Cognitive Brain Research, 6(1), 37-52.

Pensalfini, R. (2002). Vowel harmony in Jingulu. Lingua, 112(7), 561-586.

Peperkamp, S., Dupoux, E., & Sebastián-Gallés, N. (1999). Perception of stress by French, Spanish, and bilingual subjects. 6th European Conference on Speech Communication and Technology (Eurospeech 1999).

Perlmutter, D. (1988). The Split Morphology Hypothesis: Evidence from Yiddish. In M. Hammond & M. Noonan (Eds.), Theoretical Morphology: Approaches in Modern Linguistics (pp. 79-100). Tokyo/Toronto: Academic Press Inc. Harcourt, Brace, Jovanovich.

Peters, A. M. (1983). The Units of Language Acquisition. Cambridge University Press.

Peters, A. M. (1997). Language typology, prosody, and the acquisition of grammatical morphemes. The crosslinguistic study of language acquisition, 5, 135-197.

Peters, A. M. (2001). Filler syllables: what is the status in emerging grammar? Journal of Child Language, 28, 229-242.

Petitto, L. A., & Marentette, P. F. (1991). Babbling in the manual mode: Evidence for the ontogeny of language. Science, 251(5000), 1493-1496.

Phillips, B. S. (2001). Lexical diffusion, lexical frequency, and lexical analysis. In J. Bybee & P. Hopper (Eds.), Frequency and the emergence of linguistic structure (pp. 123-136). Amsterdam: Benjamins.

Phoon, H. S., Abdullah, A. C., Lee, L. W., & Murugaiah, P. (2014). Consonant acquisition in the Malay language: A cross-sectional study of preschool aged Malay children. Clinical linguistics & phonetics, 28(5), 329-345.

Pinker, S. (1989). Learnability and Cognition: The Acquisition of Argument Structure. Cambridge, MA: MIT Press.

Pinker, S. (1991). Rules of language. Science, 253(5019), 530-535.

Pinker, S. (1995). Facts about human language relevant to its evolution. In J. P. Changeux & J. Chavaillon (Eds.), Origins of the human brain (pp. 262-285). New York: Oxford University Press.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. Cognition, 28(1-2), 73-193.

Pinker, S., & Prince, A. (1991). Regular and irregular morphology and the psychological status of rules of grammar. Berkeley Linguistics Society, 17, 230-251.

Plank, F. (1994). Inflection and Derivation. In R. E. Asher (Ed.), The Encyclopedia of Language and Linguistics, Vol III (pp. 1671-1678). Oxford: Pergamon Press.

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. Cognition, 48(1), 21-69.

Prasada, S., & Pinker, S. (1993). Generalization of regular and irregular morphological patterns. Language and cognitive processes, 8(1), 1-56.

Prince, A., & Smolensky, P. (1993). Optimality Theory: Constraint interaction in generative grammar. Manuscript, Rutgers University and University of Colorado, Boulder.

Prunet, J. F., Béland, R., & Idrissi, A. (2000). The mental representation of Semitic words. Linguistic inquiry, 31(4), 609-648.

Pulvermüller, F., Härle, M., & Hummel, F. (2001). Walking or talking?: Behavioral and neurophysiological correlates of action verb processing. Brain and language, 78(2), 143-168.

Ramat, P. (2011). Linguistic Typology. In Linguistic Typology. De Gruyter Mouton.

Rastle, K., Lavric, A., Elchlepp, H., & Crepaldi, D. (2015). Processing differences across regular and irregular inflections revealed through ERPs. Journal of Experimental Psychology: Human Perception and Performance, 41(3), 747.

Ravid, D. (2003). A developmental perspective on root perception in Hebrew and Palestinian Arabic. Language Acquisition and Language Disorders, 28, 293-320.

Reesink, G. P. (1999). A Grammar of Hatam: Bird's Head Peninsula, Irian Jaya. Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In Speech and Language (Vol. 10, pp. 243-335). Elsevier.

Revithiadou, A. (1998). Lexical marking and dominance in Modern Greek. Themes in Greek Linguistics, 23-50.

Rice, K. (2000). Morpheme order and semantic scope: Word formation in the Athapaskan verb (Cambridge Studies in Linguistics 90). Cambridge: Cambridge University Press.

Rice, K. (2005). "Prominence and the verb stem in Slave (Hare)," in Hargus, S. and Rice, K. (eds.), Athabaskan Prosody. Amsterdam/Philadelphia: John Benjamins.

Rijkhoff, J., & Bakker, D. (1998). Language sampling. Linguistic Typology, 2, 263-314.

Rivera, D. M. S. (2013). Formación e historia de vuecencia en español como proceso de rutinización lingüística. Iberoromania, 77(1), 108-129.

Robbeets, M., & Savelyev, A. (2020). The Oxford Guide to the Transeurasian Languages. Oxford University Press.

Robins, R. H. (1958). The Yurok Language (Vol. 15). Berkeley and Los Angeles: University of California Press.

Robinson, S. (2011). Split intransitivity in Rotokas, a Papuan language of Bougainville (Doctoral Dissertation). Radboud University Nijmegen.

Rodriguez-Fornells, A., Clahsen, H., Lleo, C., Zaake, W., & Münte, T. F. (2001). Event-related brain responses to morphological violations in Catalan. Cognitive Brain Research, 11(1), 47-58.

Romero-Méndez, R. (2009). A reference grammar of Ayutla Mixe (Tukyo'm ayuujk) (Doctoral dissertation). State University of New York at Buffalo.

Rose, R. L. (2005). The relative contribution of syntactic and semantic prominence to the salience of discourse entities. Dissertation submitted at Northwestern University, Evanston, Illinois.

Royle, P. (2007). Variable effects of morphology and frequency on inflection patterns in French preschoolers. *The Mental Lexicon*, *2*(1), 103-125.

Royle, P., Beritognolo, G., & Bergeron, E. (2012). Regularity, sub-regularity and irregularity in French acquisition. In J. van er Auwera, T. SToz, A. Urdze, & H. Otsuka (Eds.), Irregularity in Morphology (and beyond) (pp. 227-250). Berlin: Akademie Verlag.

Sagarra, N. (2001). On L2 oral comprehension. In Romance Syntax, Semantics and L2 Acquisition: Selected Papers from the 30th Linguistic Symposium on Romance Languages, Gainesville, Florida, February 2000 (Vol. 216, pp. 197-210). John Benjamins Publishing.

Samek-Lodovici, V. (2005). Prosody-syntax interaction in the expression of focus. NLLT 23, 687-755.

Sampson, G., Gil, D., & Trudgill, P. (Eds.). (2009). Language complexity as an evolving variable (Vol. 13). Oxford University Press.

Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. Cognition, 69(2), 105-134.

Sapir, Edward (1921). Language: An introduction to the study of speech. New York: Harcourt, Brace & Co.

Saville-Troike, M. (1996). Development of the Inflected Verb in Navajo Child language. In Jelinek, E., Midgette, S., Rice, K. (Eds.), Athabaskan Language Studies: Essays in Honor of Robert W. Young (pp. 137-192). University of New Mexico Press.

Say, T., & Clahsen, H. (2002). Words, rules and stems in the Italian mental lexicon. In S. Nooteboom, F. Weerman & F. Wijnen (Eds.), Storage and Computation in the Language Faculty (pp. 93-129). Kluwer: Dordrecht.

Schmid, H. J. (2000). English Abstract Nouns as Conceptual Shells. From corpus to cognition. Berlin / New York: Mouton de Gruyter.

Schmid, H. J. (2007). "Entrenchment, salience, and basic levels," in Geeraerts, D. and Cuyckens, H. (Eds.), The Oxford handbook of cognitive linguistics (pp. 117-138).

Schmid, H. J. (2010). Does frequency in text instantiate entrenchment in the cognitive system. Quantitative methods in cognitive semantics: Corpus-driven approaches, 101-133.

Schmid, H. J. (2020). The dynamics of the linguistic system: Usage, conventionalization, and entrenchment. Oxford University Press.

Schoenemann, P. T. (1999). Syntax as an emergent characteristic of the evolution of semantic complexity. Minds and Machines, 9, 309- 346.

Ségéral, P., & Scheer, T. (2008). "The Coda Mirror, stress and positional parameters," in De Carvalho, J. B., Scheer T., Ségéral P. Lenition and fortition (pp. 483-518). De Gruyter Mouton.

Seifart, F., et al. (2021). The extent and degree of utterance-final word lengthening in spontaneous speech from 10 languages. Linguistics Vanguard, 7(1).

Selkirk, E. (1995). "Sentence prosody: intonation, stress, and phrasing," in Goldsmith, J. A. (ed.), Handbook of Phonological Theory (pp. 550-569). Cambridge, Massachusetts: Blackwell.

Sidwell, P. (2000). Proto South Bahnaric: A Reconstruction of a Mon-Khmer Language of Indo China. Canberra: Pacific Linguistics.

Sinnemäki K. (2009). Complexity in core argument marking and population size. In: Language complexity as an evolving variable (pp. 126-140). Oxford University Press.

Sinnemäki, K. (2014). Cognitive processing, language typology, and variation. Wiley Interdisciplinary Reviews: Cognitive Science, 5(4), 477-487.

Skirgård, Hedving et al. (2023). Grambank v1.0 (v1.0.3) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7844558

Skoruppa, K., Lambrechts, A., & Peperkamp, S. (2011). The role of phonetic distance in the acquisition of phonological alternations. In Proceedings of the 39th North Eastern Linguistics Conference (pp. 717-729). Somerville, MA: Cascadilla Press.

Slobin, D. I. (1973). "Cognitive Prerequisites for the Development of Grammar," in Ferguson, C. and Slobin, D. I. (eds.), Studies of Child Language Development (pp. 175-208). New York: Holt, Rinehart and Winston.

Slobin, D. I. (1985) (ed.). The crosslinguistic study of language acquisition. New Jersey: Lawrence Erlbaum.

Smeets, I. (2008). A grammar of Mapuche. De Gruyter Mouton.

Smith, C. L., Erickson, D., & Savariaux, C. (2019). Articulatory and acoustic correlates of prominence in French: Comparing L1 and L2 speakers. Journal of Phonetics, 77, 100938.

Smits, H. J. (2017). A grammar of Lumun. A Kordofanian Language of Sudan. Cologne: Rüdiger Köppe.

Smolka, E., & Khader, P. H. (2013). Electrophysiological evidence for the continuous processing of linguistic categories of regular and irregular verb inflection in German. Journal of Cognitive Neuroscience, 25(8), 1284-1304.

Smolka, E., & Eulitz, C. (2018). Psycholinguistic measures for German verb pairs: Semantic transparency, semantic relatedness, verb family size, and age of reading acquisition. Behavior research methods, 50(4), 1540-1562.

Sohn, H. M. (1999). The Korean language. Cambridge University Press.

Solomonoff, R. J. (1964). A formal theory of inductive inference. Part I. Information and control, 7(1), 1-22.

Song, J. (2013). Variable schwa deletion in English: A corpus study. 음성음운형태론연구, 19(1), 33-52. Stanford University Press.

Steels, L. (2000, September). Language as a complex adaptive system. In International Conference on Parallel Problem Solving from Nature (pp. 17-26). Springer, Berlin, Heidelberg.

Stockall, L., & Marantz, A. (2006). A single route, full decomposition model of morphological complexity: MEG evidence. The mental lexicon, 1(1), 85-123.

Stump, G. (2019). Some sources of apparent gaps in derivational paradigms. Morphology, 29(2), 271-292.

Swadesh, M. (1952). Lexicostatistic Dating of Prehistoric Ethnic Contacts. Proceedings of the American Philosophical Society, 96, 452-463.

Szendroï, K. (2003). A stress-based approach to the syntax of Hungarian focus. The Linguistic Review, 20, 37-78.

Szmrecsanyi, B. (2004, March). On operationalizing syntactic complexity. In Le poids des mots. Proceedings of the 7th international conference on textual data statistical analysis. Louvain-la-Neuve (Vol. 2, pp. 1032-1039).

Taatgen, N., & Dijkstra, M. (2003). Constraints on Generalization: Why are Past-Tense Irregularization Errors so Rare?. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 25, No. 25). Mahwah, NJ: Erlbaum.

Tabain, M. (2003). Effects of prosodic boundary on /aC/ sequences: Articulatory results. Journal of the Acoustical Society of America, 113, 2834-2849.

Tardif, T., (2005, April). But are they really verbs? Paper presented at the meeting of the Society for Child Development, Atlanta, Georgia.

Temperley, Nicholas and David Temperley. Stress-meter alignment in French vocal music. The Journal of the Acoustical Society of America, 134(1), 520-527.

Terken, J. and Hermes, D. (2000). "The perception of prosodic prominence," in M. Horne (ed.), Prosody: Theory and Experiment, 89-127. Dordrecht: Kluwer.

Thomason, S. G. (1980). "Morphological instability, with and without language contact," in Fisiak, J. (ed.), Historical morphology, 359-372, Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110823127.359

Thomason, S. G. (2001). Language contact. Edinburgh University Press.

Tkachman, O., Hall, K. C., Fuhrman, R., & Aonuki, Y. (2019). Visible amplitude: Towards quantifying prominence in sign language. Journal of Phonetics, 77, 100935.

Todd, S., Pierrehumbert, J. B., & Hay, J. (2019). Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. Cognition, 185, 1-20.

Tomasello, M. (2000). Do young children have adult syntactic competence? Cognition, 74(3), 209-253.

Traugott, E. C. (2002). From etymology to historical pragmatics. Studies in the history of the English language: A millennial perspective, 39, 19-50.

Traugott, E. C., & Dasher, R. B. (2001). Regularity in Semantic Change. Cambridge University Press, Cambridge Studies in Linguistics 96.

Traugott, E. C., & Trousdale, G. (2013). Constructionalization and constructional changes (Vol. 6). OUP Oxford.

Trudgill, P. (1999). Language contact and the function of linguistic gender. Poznan Studies in Contemporary Linguistics, 35, 133-152.

Trudgill, P. (2001). Contact and simplification: Historical baggage and directionality in linguistic change. Linguistic Typology, 5(2/3), 371-374.

Trudgill, P. (2011). Sociolinguistic Typology. Social determinants of linguistic typology. Oxford University Press.

Tsao, F. M. (2008). The effect of acoustical similarity on lexical-tone perception of one-year-old Mandarin-learning infants. Chinese Journal of Psychology, 50, 111-124. doi: 10.6129/CJP.2008.5002.01

Tsao, F. M. (2017). Perceptual improvement of lexical tones in infants: effects of tone language experience. Frontiers in psychology, 8, 558. doi: 10.3389/fpsyg.2017.00558

Tsuchida, A., & Recasens, D. (1998). Phonetics and phonology of Japanese vowel devoicing. Glot international, 3(9-10), 10-11.

Tucker, A. N., & Creider, C. A. (1994). A Grammar of Kenya Luo (Dholuo) (Vol. 2). Köln: Rüdiger Köppe.

Tucker, W. (1969). Chronology of Greek Sound Changes. The American Journal of Philology, 90(1), 36-47.

Ullman, M. T. (2001). The declarative/procedural model of lexicon and grammar. Journal of psycholinguistic research, 30(1), 37-69.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. Cognition, 92(1-2), 231-270.

Ullman, M. T., Pancheva, R., Love, T., Yee, E., Swinney, D., & Hickok, G. (2005). Neural correlates of lexicon and grammar: Evidence from the production, reading, and judgment of inflection in aphasia. Brain and Language, 93(2), 185-238.

Vaissière, J. (2002). "Cross-linguistic prosodic transcription: French versus English," in Volslkaya, N. B., Svetozarova, N. D., & Skrelin, P. A. (eds.), Problems and methods in experimental phonetics, In honour of the 70th anniversary of Prof. L.V. Bondarko, 147-164. St. Petersburg: St. Petersburg State University.

Vajda E. J. (2003). Tone and phoneme in Ket. Amsterdam Studies in the Theory and History of Linguistic Science Series 4: 393-418

Vajda E. J. (2004). Ket (Languages of the world: Materials 204). Lincom Europa, Munich

Vajda, E. (2010). A Siberian link with Na-Dene languages (Vol. 5, pp. 33-99). Anthropological Papers of the University of Alaska: New Series.

Vajda, E. (2019). Morphology in Dene-Yeniseian Languages. In Oxford Research Encyclopedia of Linguistics.

Van Driem, G. (2011). A grammar of Dumi (Vol. 10). Berlin/New York: Walter de Gruyter.

Van Geert, P. (2008). The dynamic systems approach in the study of L1 and L2 acquisition: An introduction. The Modern Language Journal, 92(2), 179-199.

Van Geert, P., & Verspoor, M. (2015). Dynamic systems and language development. In The handbook of language emergence.

Varden, J. (1998). On High Vowel Devoicing in Standard Modern Japanese: Implications for Modern Phonological Theory (Doctoral Dissertation). University of Washington.

Veneziano, E. (1988). "Vocal-verbal interaction and the construction of early lexical knowledge," in Smith, M.D. and Locke, J. L. (eds.) The emergent lexicon: the child's development of a linguistic vocabulary, 109-147. New York, NY: Academic Press.

Verluyten, P. (1984). Phonetic reality of linguistic structures: The case of (secondary) stress in French. In Proceedings of the 10th International Congress of Phonetic Sciences, Utrecht, 522-526.

Vidal, A. (2001). Pilagá grammar (Guaykuruan family, Argentina) (Doctoral Dissertation). University of Oregon.

Vigil, N., & Oller, J. (1976). Rule fossilization: A tentative model. Language Learning, 26, 148—162.

Vihman, M., & Croft, W. (2007). Phonological development: toward a "radical" templatic phonology. Linguistics, 45(4), 683-725. https://doi.org/10.1515/LING.2007.021

Vihman, M. M. (2010). Phonological templates in early words. In Fougeron, C., Kühnert, B., D'Imperio, M., Vallée, N. Laboratory phonology. 261-284. Berlin/New York: Walter de Gruyter.

Villard, S. (2015). The phonology and morphology of Zacatepec Eastern Chatino (Doctoral dissertation). University of Texas at Austin.

Wagner, S., Smith, K., & Culbertson, J. (2019). Acquiring agglutinating and fusional languages can be similarly difficult: Evidence from an Adaptive Tracking Study. In A.K. Goel, C.M. Seifert, & Freksa (Eds.) Proceedings of the 41st Annual Conference of the Cognitive Science Society. Montreal, QB: Cognitive Science Society, 3050-3056.

Walsh Dickey, L. (1999). Syllable count and Tzeltal segmental allomorphy. In Phonological 1996: Syllables!?, ed. by John R. Rennison and Klaus Kühnhammer, The Hague: Holland Academic Graphics, 323-334.

Walsh, M. J. (1976). The Murinypata Language of North-West Australia. (Doctoral Dissertation). Australian National University.

Wang, P. (2015). Phonological Prominence and Its Interaction with Tone in Chinese Dialects (Doctoral Dissertation). University of Pittsburgh.

Waterson, N. (1971). Child phonology. Journal of Linguistics, 7, 179-211.

Watkins, L. J. (1980). A grammar of Kiowa. University of Kansas.

Watkins, L. J. (1984). A Grammar of Kiowa. University of Nebraska Press.

Watson, K. (2006). Lenition and segmental interaction: evidence from Liverpool English (and Spanish). Glossa, I, 1.

Werner H. (2002). Vergleichendes Wörterbuch der Jenissej-Sprachen, Vol. 1-3. Otto Wiesbaden: Harrassowitz Verlag.

Weyerts, H., Münte, T. F., Smid, H. G., & Heinze, H. J. (1996). Mental representations of morphologically complex words: an event-related potential study with adult humans. Neuroscience Letters, 206(2-3), 125-128.

Whiting, C. M., Marslen-Wilson, W. D., & Shtyrov, Y. (2013). Neural dynamics of inflectional and derivational processing in spoken word comprehension: laterality and automaticity. Frontiers in human neuroscience, 7, 759.

Wilcox, S., & Occhino, C. (2016). Constructing signs: Place as a symbolic structure in signed languages. Cognitive Linguistics, 27(3), 371-404.

Wit, E. J. C., & Gillette, M. (1999). What is linguistic redundancy? University of Chicago.

Xu, F., & Pinker, S. (1995). Weird past tense forms. Journal of child language, 22(3), 531-556.

Yadav, H., Vaidya, A., Shukla, V., & Husain, S. (2020). Word Order Typology Interacts with Linguistic Complexity: A Cross-Linguistic Corpus Study. Cognitive science, 44(4), e12822.

Yip, M. (2001). The complex interaction of tone and prominence. In North East Linguistics Society (Vol. 31, No. 2), 19.

Young, R. W., Morgan, W., & Midgette, S. (1992). Analytical lexicon of Navajo. University of New Mexico Press.

Yue-Hashimoto, A. O. (1986). Tonal flip-flop in Chinese dialects. Journal of Chinese Linguistics, 14(2), 161-183.

Zamponi, R. (2003). Betoi (Languages of the World/Materials 428). München: Lincom Europa.

Zepeda, O. (1983). A Tohono O’odham grammar. University of Arizona Press.

Zimmermann, M. (2016). Predicate focus. In: Féry C. and Ishihara S.: The Oxford handbook of information structure. Retrieved from DOI:10.1093/oxfordhb/9780199642670.001.0001

Zingler, T. (2020). Wordhood issues: typology and grammaticalization (Doctoral dissertation). The University of New Mexico.

Zipf, G. K. (1935). Psycho-Biology of Language. Boston: Houghton Mifflin.

Zipf, G. K. (1949). Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology. Cambridge, Massachusetts: Addison-Wesley.

Zúñiga, F. (2006). Deixis and Alignment: Inverse systems in indigenous languages of the Americas (Vol. 70). John Benjamins Publishing.