5-1-2016

Illustrating the prototype structures of parts of speech: A multidimensional scaling analysis

Phillip Rogers

Phillip Gordon Rogers
Candidate

Linguistics
Department

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

Approved by the Thesis Committee:

William Croft, Chairperson

Melissa Axelrod

Ian Maddieson

ILLUSTRATING THE PROTOTYPE STRUCTURES OF PARTS OF SPEECH: A MULTIDIMENSIONAL SCALING ANALYSIS

by

PHILLIP GORDON ROGERS

B.A., HISTORY, OHIO DOMINICAN UNIVERSITY, 2009

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Arts
Linguistics

The University of New Mexico
Albuquerque, New Mexico

May 2016

ACKNOWLEDGMENTS

First, I want to thank my advisor, Dr. Bill Croft, for intellectual inspiration, thoughtful guidance and encouragement, and remarkable patience. His influence permeates all aspects of this thesis.

I am also grateful for the support of committee members Dr. Melissa Axelrod and Dr. Ian Maddieson. They have never hesitated to offer their time and wisdom—even on short notice—regarding this thesis or other research.

A special thanks goes out to Dr. James and Maria Mondloch for their time and willingness to share their knowledge of K'ichee' Mayan. Even in the face of abstract and relatively unnatural elicitation prompts, they responded with patience and valuable insight.

I would like to acknowledge friends and colleagues with whom I have discussed this research or related topics, including Daniel Hieber, Matthew Menzenski, Stefano de Pascale, Mitchell Sances, Logan Sutton, Michael Woods, and Tim Zingler.

Last, and most importantly, I would like to thank my wife Christa, my children—Isaac, NoaRose, and Micah—and the rest of our families. Time spent on this thesis was time away from them. In particular, Christa has been both the heart and the cornerstone of our household, keeping me grounded and our kids inspired.

ILLUSTRATING THE PROTOTYPE STRUCTURES OF PARTS OF SPEECH: A MULTIDIMENSIONAL SCALING ANALYSIS

by

Phillip Gordon Rogers

B.A., History, Ohio Dominican University, 2009
M.A., Linguistics, University of New Mexico, 2016

ABSTRACT

Radical Construction Grammar (Croft 2001) proposes that parts of speech can be explained as prototypes that emerge from the use of broad semantic classes of words—objects, properties, and actions—in basic propositional act functions of discourse—reference, modification, and predication. This theory predicts that each of these broad semantic classes will be typologically unmarked in its prototypical propositional act function and relatively marked in other propositional act functions.

Because this theory speaks to such a broad and fundamental organization of linguistic structure, the rich structure of these prototypes has not been fully explored in a comprehensive manner. Gradience is a key characteristic of prototypes (Rosch 1978), and it is found for parts of speech in the continuum from object concepts to property concepts to action concepts. This gradience—and the semantic primitives that motivate the continuum—are explored in more detail within each of these broad semantic classes through a discussion of the literature on noun classes, adjective classes, and verb classes.

Equipped with a list of conceptual targets that are predicted to represent the range of prototypicality for each part of speech, this thesis sets out to illustrate their prototype structures. For eleven genetically, geographically, and typologically diverse languages, lexemes are identified to represent these conceptual targets. The criteria of typological markedness is used to identify asymmetries in how these lexemes are formally encoded relative to each other across the three propositional act functions. These markedness asymmetries are then coded for and illustrated in a multidimensional scaling analysis.

The insight of the spatial model is twofold: First, it illustrates the true prototype structures of parts of speech in which many concepts cluster at the prototypes and a few end up on the peripheries. Even more, it reveals the relative strengths of the prototypes, confirming the hypothesis of a weaker adjective prototype. Second, the spatial model sheds light on the semantic primitives that motivate the prototype structures, both internally and in relation to each other.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

1	first person
2	second person
3	third person
ACC	accusative
ASSOC	associative
B	noun class “B” marker (Ingush)
CAUS	causative
CONTR	contrastive
COP	copula
DEM	demonstrative
DENOM	denominalizer
DUR	durative
FOC	focus
FUT	future
III	noun class “III” marker (Mali)
INF	infinitive
IPFV	imperfective
LOC	locative
M	masculine
NFUT	non-future
NPRS	non-present
NPST	non-past
OBJ	object
PERS	personal
PL	plural
POSS	possessive
PRED	predicating particle
PRF	perfect
PRS	present
PST	past
RDP	reduplication
SG	singular
V	noun class “V” marker (Ingush)
VENT	venitive

1. Introduction

Radical Construction Grammar (Croft 2001) proposes that parts of speech (noun, adjective, and verb)[1] can be explained as prototypes that emerge from the use of broad semantic classes of words—objects, properties, and actions—in basic propositional act functions of discourse—reference, modification, and predication. This theory predicts that each of these broad semantic classes will be typologically unmarked in its prototypical propositional act function and relatively marked in other propositional act functions.

Because this theory addresses such a broad and fundamental organization of linguistic structure, the rich structure of these prototypes has not been fully explored in a comprehensive manner. Gradience is a key characteristic of prototypes (Rosch 1978), and it is found for parts of speech in a conceptual continuum from objects to properties to actions. This gradience will be explored in more detail within each of these broad semantic classes through a discussion of the literature on noun classes, adjective classes, and verb classes.

[1] Certain labeling conventions are used throughout this paper to make clear the distinction between universal categories and language-particular ones (following Comrie 1976; Bybee 1985; Croft 2001). Semantic notions such as properties or semelfactives, as well as propositional act functions such as predication, will be written with a lowercase first letter. These are cross-linguistic concepts. Language-particular categories, such as Adjectives or Participles, are capitalized. Labels for reoccurring morphosyntactic strategies, such as relative clauses or predicate nominals, are lowercase, even if there are languages that do not use that particular strategy. Finally, the categories Noun, Adjective, and Verb are often capitalized, following the theoretical stance of this paper that they represent language-particular categories. However, they do appear in lowercase when they refer—for sake of convenience—to the conceptual-functional prototypes, even for previous research which implicitly or explicitly characterizes them differently.

Equipped with a list of conceptual targets that are predicted to represent the range of prototypicality for each of the parts of speech, this thesis sets out to illustrate these prototype structures. For eleven genetically, geographically, and typologically diverse languages, lexemes will be identified to represent these conceptual targets. I will use the criteria of typological markedness to identify asymmetries in how these lexemes are formally encoded relative to each other across the three propositional act functions. These markedness asymmetries will be coded for and illustrated in a multidimensional scaling analysis.

The goals of this multidimensional scaling analysis are twofold: First, the analysis should confirm the characteristics of true prototype structures: many concepts clustered at the prototypes and a few concepts inhabiting the peripheries—particularly between two prototypes. Even more, the relative strengths of the prototypes should be evident in the analysis. Of particular interest is the adjective prototype, which has been claimed to be weaker or even incomparable to that of nouns and verbs. Second, the multidimensional scaling analysis can shed light on the semantic primitives that motivate the prototype structures, as relevant semantic factors will correlate strongly with particular dimensions in the spatial model.

2. Background

The exact nature of parts of speech has been a central topic of linguistic inquiry and debate since at least antiquity. Long before native languages of the Pacific Northwest were presented as a challenge to the universal distinction between nouns and verbs, Plato famously disputed the reality of 'tableness' and 'cupness' with Diogenes the Cynic (Stanley 1687). At stake is a fundamental organization of language structure, and many terms have been used for this phenomenon with greater or lesser theoretical implications: parts of speech, lexical classes, word classes, grammatical categories, etc. In this paper I will most often use parts of speech for the sake of consistency; when any of the other terms are used, they are meant to be synonymous. Furthermore, parts of speech can generally refer to all the word classes in a language, from large, lexical, open classes to small, functional, closed classes—and everything in between. This thesis focuses on the basic parts of speech commonly referred to as nouns, verbs, and adjectives. Adverbs are often included with these three, and although they can be explained by the theory to which this paper ascribes, they are excluded in this study in order to reasonably limit its scope.

It is a widely-held belief that parts of speech should be identified in a particular language using formal criteria (Schachter & Shopen 2007). These criteria include a word's syntactic distribution, its syntactic function, and its ability to take various syntactic and/or morphological inflections. It follows from this that parts of speech are language-specific, and there are countless examples in the literature to illustrate how a particular lexical class in one language does not have exactly the same members as the purportedly equivalent lexical class in another language. In fact, the various criteria *within* a language do not always yield the same classifications, and this leads to important theoretical questions (Schachter & Shopen 2007: 4): Are there truly discrete distinctions among parts of speech? When does a cluster of grammatical criteria indicate a distinction between classes, and when does it indicate a distinction between subclasses? Which criteria are most important? Most analyses of parts of speech in particular languages fail to answer these questions explicitly, resulting in some amount of arbitrariness. It seems that formal criteria are not sufficient to explain both the language-particular aspects and cross-linguistic similarities of parts of speech.

Although the delineation of parts of speech in a language is accomplished on formal grounds, the terminology applied to a lexical class once it has been identified is often based on the semantic nature of its members. For example, a lexical class identified in a particular language will likely be labeled as Nouns if most of its members include people, places, and things. However, this method for matching language-particular word classes with unifying semantic labels is not always straightforward. In a language with two distinct classes of property words, which one gets the traditional Adjective label and what label is used for the other class? These kinds of semantic considerations are not enough to validate comparison of Adjectives in one language to Adjectives in another. The fact is that languages make different decisions regarding how many lexical classes, where to draw the line between them, and by which formal means.

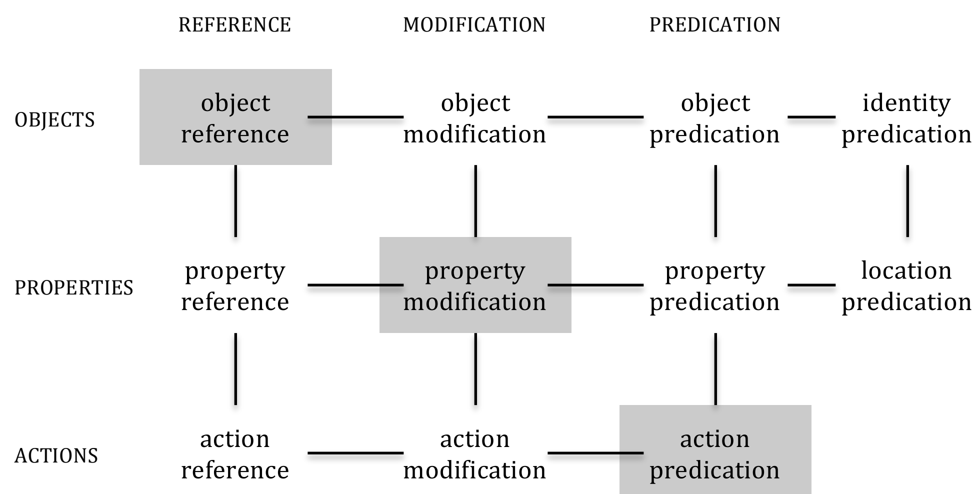
The missing component in our understanding of parts of speech thus far is discourse function. Hopper and Thompson (1984) argue that all linguistic forms lack 'categoriality' until it is forced on them by use in a particular discourse function. However—foreshadowing a prototype theory that invokes both discourse function and lexical semantics—they write that "most forms begin with a propensity or predisposition to become Nouns or Verbs" (1984: 747). More recent models (Croft 1991; Hengeveld 1992) invoke the propositional act functions of reference, modification, and predication, which are based on Searle's (1969) work on speech acts. These propositional act functions represent distinct communicative goals: reference serves to establish a referent, modification to modify an existing referent, and predication to report on a referent's state of affairs.

2.1. Parts of speech as prototypes

Bringing together previous insights, Croft (1991, 2001) proposes that parts of speech can be explained as prototypes that emerge from the use of broad semantic classes of words—objects, properties, and actions—in the basic propositional act functions of discourse—reference, modification, and predication. This theory predicts that each of these broad semantic classes will be typologically unmarked in its prototypical propositional act function—objects in reference, properties in modification, and actions in predication—and relatively marked in other propositional act functions. Typological markedness will be defined and discussed in more detail in section 2.3, and it plays an important role in the methodology of this paper. For now it will suffice to say that patterns of typological markedness are based on formal criteria, which as we have seen is a critical component of any theory of parts of speech.

Moreover, the prototype theory of parts of speech also accounts for both the semantic nature of parts of speech and the role of discourse function. Figure 1 features a conceptual space (this tool is discussed in more detail in section 2.5.1) for parts of speech that illustrates how each broad semantic class can be employed in each propositional act function. The darker sections are the prototypes: each broad semantic class is a prototype of the constructions of one propositional act function.

Figure 1. Conceptual space for parts of speech (Croft 2001: 92)



2.2. Prototype structure and the semantic continuum of parts of speech

Croft's conceptual space for parts of speech illustrates well the prototypes that result from combinations of broad semantic class and propositional act function. However, a fundamental characteristic of prototype structure is the absence of discrete boundaries (Rosch 1978). Instead, prototypes have a gradient dimension along which members of the category can be identified as more or less prototypical to the category. This gradience arises from the intersection of many individual attributes, which in turn can be either discrete or gradient. The more prototypical a member is of a category, the more attributes it has in common with other members of the category and the fewer attributes it has in common with members of contrasting categories.

For parts of speech, the continuum from objects to properties to actions is a gradient dimension. Concerning the conceptual space of parts of speech, Croft writes that, "the vertical dimension could be as finely divided as single word concepts if necessary" (Croft 2001: 94). This continuum hypothesis was first developed by Ross (1972), and has been refined by several others since (e.g., Comrie 1975; Givón 1979; Dixon 1982; Croft 1991, 2001). Givón (1979) introduced the idea that time-stability differentiates nouns from verbs, and that adjectives fall somewhere in the middle of this continuum. Croft (2001) splits this concept into transitoriness and stativity, and he includes two others: relationality and gradability. A hypothesis of this paper is that the multidimensional scaling analysis will reflect the influence of these semantic primitives. More specifically, they are expected to differentiate object, property, and action concept clusters.

Table 1. Semantic properties of prototypical parts of speech (Croft 2001: 87)

	Relationality	*Stativity*	*Transitoriness*	*Gradability*
Objects	nonrelational	state	permanent	nongradable
Properties	relational	state	permanent	gradable
Actions	relational	process	transitory	nongradable

The semantic primitives in Table 1 were proposed to motivate the distinctions *among* the three broad semantic classes. But what about the internal structure of each? It is clear from the literature that these classes are not internally homogeneous either formally or semantically. Fortunately, much has been written about the formal and semantic variability within these groups, although mostly under the terminology of noun, adjective, and verb classes. In fact, different semantic factors seem to be relevant to variability and prototypicality within each of these classes. For each one in turn, I will briefly review the pertinent literature and use the insights therein to build a list of conceptual targets for use in this study. The goal is a list of conceptual targets that represent the full range of prototypicality on whichever semantic dimensions are expected to be relevant. If the semantic factors discussed in the literature are truly motivating prototypicality for a particular part of speech, the prototype structure will be borne out in the spatial model.

2.2.1. Object subclassification

The most important development for understanding object concept prototypicality has been the extended animacy hierarchy first formulated by Silverstein (1976; see also Dixon 1979; Comrie 1989; Croft 2003). This hierarchy actually includes three semantic dimensions: person, referentiality, and animacy (Croft 2003: 130). The role of the extended animacy hierarchy can be found in many areas of morphosyntax. A well-known example is the inflection of nouns for number (Corbett 2000). If a language inflects animate nouns for number, it will also inflect for number all the categories that are higher on hierarchy (e.g., human nouns and pronouns), but not necessarily those lower on the hierarchy (e.g., inanimates). Some languages inflect most or all pronouns/nouns for number, while others inflect only a very limited subset—either way, the hierarchy determines which categories are included in that subset. It has been suggested that these semantic factors interact with other factors such as definiteness and topicality (Comrie 1989); this is almost certainly true, but it is still unclear how much of their effect on morphosyntactic patterns is direct and how much is indirect.

The role of person is found among pronouns, where first and second person pronouns generally outrank third person pronouns. For purposes of scope and in consideration of the methodological challenges that they would present, I chose not to include pronouns in this study. There is evidence that pronouns are "superprototypical" members of the noun prototype, less marked than even the most prototypical common nouns (Croft 1991: 126-8). For example, Toqabaqita Pronouns make a three-way number distinction while Nouns contrast only singular and plural (Lichtenberk 2008: 325). Similarly, referentiality is relevant in distinguishing pronouns from proper nouns from common nouns. Like pronouns, proper nouns were left out of this study, leaving no variation upon which referentiality might act.

This leaves animacy, which is hypothesized for this study to motivate prototypicality among object concepts. The list of object concepts in Table 2 is based on well-known noun classes (see for example Lehmann & Moravcsik 2000; Croft, to appear), and it strongly reflects a dimension of animacy. For each class, exemplars were chosen for their likelihood to be found in all or most of the languages of this study. Where it was possible to include two exemplars from a single class, complements such as male/female, higher animate/lower animate, large/small, manmade/natural, etc. were included to maximize variation. The same principles for choosing exemplars were applied for the property concept and action concept subclassifications discussed below.

Table 2. Object concept subclassification and selected exemplars

Class	***Exemplars***
Humans	WOMAN
	BOY
Kinship	FATHER
	SISTER
Animals	DOG
	BIRD
Plants/Plant Products	TREE
	SEED
Artifacts	HAT
	BED
Body Parts	ARM
	FACE
Places	RIVER
	HOUSE
Material/Substance	WATER
	STONE

2.2.2. Property subclassification

Adjectives (and the property concepts they prototypically encode) have received a great deal of attention over the years for their controversial part of speech status. Some linguists assert that adjectives are universal and on par with nouns and verbs as basic parts of speech, while others cite evidence from particular languages in which adjectives appear to be incomparable or non-existent. While this study does find evidence for the validity of a universal adjective *prototype*, this finding does not necessitate that every language has a solitary, dedicated, and easily delineated category of Adjectives. Rather, it is clear that language-particular Adjective categories come in different shapes and sizes, so to speak.

While the variation in adjective classes is substantial, it is not unconstrained. In his famous article/chapter, “Where have all the adjectives gone?”, Dixon (1982) identified subclasses of property concepts based on their semantic characteristics. He demonstrated with cross-linguistic data how some of these subclasses are more central to the cross-linguistic notion of adjective, showing up as Adjectives even in languages with small Adjective word classes; these are age, dimension, value, and color concepts. Others were more peripheral, encoded as Adjectives only in languages with large Adjective classes. Furthermore, he found tendencies among these peripheral subclasses: physical property and speed concepts are more likely to show up as Verbs in a language (if not Adjectives), while human propensity were more likely to be Nouns. Dixon’s findings have since been confirmed and his subclassification refined by subsequent research.

There are two studies of adjectival predication that were particularly important for our understanding of property concepts. In his typology of adjectival predication, Wetzer (1996) found that even in languages where Adjectives are distinct from Nouns and Verbs, they show characteristics that make them either “nouny” or “verby”. He conceptualized these findings as language-specific treatment of property concepts on the noun-verb continuum, where adjectives are somewhere in the middle. Interestingly, he concludes that adjectives are verby by default.

In a similar vein, Leon Stassen’s *Intransitive Predication* (1997) examined formal patterns of intransitive predication in a large language sample. Semantically, he divided intransitive predication into four types: event, class-membership, locational, and property-concept (representing verbs, nouns, adverbs, and adjectives, respectively). Notably, he found that concept predicates do not have their own prototypical encoding strategy, but instead take on the strategy of one of the other types. Based on time stability, he established an adjective hierarchy similar to Dixon's, from human propensity at one end to gender and material at the other end. This continuum correlates with the frequency at which those concepts take noun-like versus verb-like encoding strategies. Furthermore, Stassen introduces a tensedness parameter that pits mandatory tense inflection on Verbs against a verby treatment of Adjectives in predication. Simply put, a language will not have both, as that would result in mandatorily assigning tense to relatively time-stable property concepts for which tense is not relevant or appropriate. The tensedness parameter supports the hypothesis that temporal semantic primitives are among those motivating the prototype structures of parts of speech.

Both Wetzer and Stassen focus on the morphosyntactic realization of properties *in predication*. As we've seen, predication is just one propositional act function in which property concepts can be used, and it isn't even the prototypical function for property concepts. However, their work refined our understanding of the where particular property concepts fall in the semantic continuum. Furthermore, their consideration of an impressive range of cross-linguistic data served to elucidate the various ways in which property concepts can be formally encoded relative to object concepts and action concepts. When we consider property concepts in propositional act functions other than predication, many of the same patterns emerge.

Just as for object concepts, eight subclasses of property concepts were identified for use in this study. These are found in Table 3, with two exemplars representing each class. The subclasses are ordered from most object-like at the top to most action-like at the bottom.

Table 3. Property concept subclassification and selected exemplars

Class	***Exemplars***
Gender	MALE
	FEMALE
Color	WHITE
	RED
Form	ROUND
	FLAT
Age	OLD
	YOUNG
Value	GOOD
	BAD
Dimension	BIG
	SHORT
Physical Properties	HEAVY
	SOFT
Human Propensity	HAPPY
	ANGRY

2.2.3. Action subclassification

The subclassification of action concepts ('verb classification', for many) has been dominated by a discussion of aspectual structure. Vendler (1967) identified four verb classes—states, activities, accomplishments, and achievements—and demonstrated for English how certain grammatical behaviors were predictable based on a Verb's aspectual classification. The distinctions among these classes can be attributed to semantic primitives such as duration, telicity, and stativity. For example, states are durative, atelic, and stative, while achievements are instantaneous, telic, and dynamic. Later research confirmed the validity of these aspectual classes for other languages, and refined the classification (Comrie 1976; Smith 1997; Croft 2012; see also Chafe 1970; Hopper & Thompson 1980; Langacker 1987; inter alia). For example, cyclic achievements (also known as semelfactives) have been identified as achievements for which the end state is identical to the beginning state. Cyclic achievements can be juxtaposed with directed achievements, in which the action results in a change of state. For example, the concept HIT can be classified as a cyclic achievement, while BREAK is classified as a directed achievement. The primary subclassification of action concepts for this study is based on these aspectual classes.

Building on the insights of Fillmore (1970), Beth Levin's (1993) comprehensive investigation of English Verbs yields a more finely-grained classification than the aspectual one described above. By identifying Verbs with similar syntactic behavior across a range of syntactic alternations, she is able to identify dozens of semantically and grammatically coherent subclasses (e.g., "prepare" verbs, "chew" verbs, etc.). Levin and others have extended these findings to include languages other than English (for discussion of cross-linguistic applicability and relevant contributions, see Levin, in press). Of course, not every one of Levin's Verb classes could be included in this study. I attempted to include exemplars from as many different Levin-inspired verb classes as possible, and in particular those classes whose members are frequent cross-linguistically.

Table 4 shows the subclassification of action concepts, with aspectual classes represented in the first column and Levin's verb classes represented in the second column.

Table 4. Action concept subclassification and selected exemplars

Class	***Type***	***Exemplars***
2-Participant States	Cognition	KNOW
	Desire	WANT
	Emotion	LOVE
Inactive Actions	Perception	SEE/LOOK AT
	Posture	SIT
	Spatial/Possession	WEAR
Durative Processes	Social Interaction	FIGHT
	Consumption	EAT
	Motion	COME
	Change of State	COOK
	Creation Verbs	BUILD
Cyclic Achievements (Semelfactives)	Emission	COUGH
	Contact	HIT
	Bodily Movement	WAVE
Directed Achievements	Existence	DIE
	Change of Possession	GIVE
	Change of State	BREAK

2.3. Typological markedness

The concept of typological markedness was invoked in the discussion of parts of speech as prototypes. In this section, I clarify the notion of markedness as it is used in this paper and discuss the criteria through which it is manifested in individual languages.

The term 'markedness' was introduced by the Prague School of linguistic theory (Trubetzkoy 1931, 1939/1969; Jakobson 1932/1984, 1939/1984, cited in Croft 2003), and has since been used in different ways by different analytical and theoretical traditions. In this paper, I refer to markedness as it applies to cross-linguistic universals, also known as 'typological markedness' (Greenberg 1966; Croft 1996). Croft (2003: 87) summarizes markedness as "asymmetric or unequal grammatical properties of otherwise equal linguistic elements: inflections, words in word classes and even paradigms of syntactic constructions." These asymmetries are manifested in particular languages through different criteria, which will be described below and exemplified in Section 3.3.

Croft makes it clear that typological markedness is "a universal property of a conceptual category" rather than the property of a specific language or construction (2003: 88). This is true for the application of markedness to parts of speech. The conceptual categories at stake are the combinations of broad semantic class and propositional act function, and as Figure 1 shows, three of these combinations are typologically unmarked relative to the others. Perhaps controversially, I will also use 'marked'/'unmarked' terminology to refer to language-specific manifestations of the universal pattern. These language-particular asymmetries are also identified by just that term—'asymmetries'—as well as their relevant criteria. However, the term 'asymmetry' does not capture the *direction* of the asymmetry, which is essential to the conception of markedness at both the language-specific and universal levels. Therefore, I will often refer to a language-particular construction as 'marked' relative to a parallel construction. It is these numerous language-specific *directional* asymmetries that collectively constitute the pattern of typological markedness for that category and its corresponding values.

Furthermore, markedness is essentially a relation between two values. A construction viewed *in isolation* cannot be described as marked or unmarked;

rather, it can be identified as such *relative* to a parallel construction according to one of several criteria. It is these criteria that I turn to next.

2.3.1. Structural coding

One criterion of typological markedness is structural coding, which entails an asymmetry in the formal expression of two or more values of a category (Croft 2003). This asymmetry is one of quantity. For structural coding, the quantity is the number of morphemes needed to express the values for the category in question[2].

> *Structural coding: the marked value of a grammatical category will be expressed by at least as many morphemes as is the unmarked value of that category* (Croft 2003: 92)

Zero coding—that is, the formal expression of a value without an overt morpheme—is a common strategy for many categories and many languages. Alternatively, values can be expressed through overt coding.

It is important to remember that the asymmetry is a universal one, which is not necessarily realized in each language. The phrasing "at least as many" in the definition above is key. Croft uses the category of number to illustrate this point. Singular is claimed to be an unmarked value for number relative to plural. Indeed, for languages in which there is an asymmetry in the number of morphemes needed to express these values, the plural is almost always the value that requires more morphemes. Yet for some languages, both values singular and plural are expressed equally in terms of number of morphemes—either they are both zero-coded, or both

[2] Counting morphemes and determining which morphemes to count are not uncontroversial endeavors. The reader is referred to Croft (2003: ch. 4) for discussion.

overtly coded. Still, these languages do not violate the claim that singular is unmarked relative to plural. Even for categories in which there is no asymmetry in a particular language, the typological markedness pattern remains valid.

2.3.2. Behavioral potential

The second criterion of typological markedness is behavioral potential, which refers to the versatility of a value relative to other values of a category. This versatility can be expressed in two ways, corresponding to two types of behavioral potential. First, inflectional potential refers to the morphological distinctions that a value can express; a value is relatively unmarked if it can inflect for more cross-cutting categories than a comparable value.

> *Inflectional potential: if the marked value has a certain number of formal distinctions in an inflectional paradigm, then the unmarked value will have at least as many formal distinctions in the same paradigm* (Croft 2003: 97)

A more subtle inflectional potential distinction can also be made based on how the inflection is accomplished morphosyntactically. From least marked to most, the inflection may be accomplished by internal change, affixation, or periphrasis. These assignments of markedness are based on token frequency (see 2.3.3): suppletive inflections are generally found on lexemes of high frequency, affixation covers a wide range of lexeme frequencies, and periphrastic constructions are often employed for low frequency lexemes.

Second, distributional potential refers to the syntactic contexts in which a value can occur; a value is relatively unmarked if it can occur in more syntactic contexts than a comparable value. Unfortunately, a comprehensive inclusion of distributional potential is beyond the scope of this study. A straightforward instance of distributional potential for parts of speech would be a language that does not allow property concepts to be used in reference, for example. However, I found no explicit evidence for this type of prohibition in any of the languages of this study.[3] There are more subtle ways in which distributional potential interacts with parts of speech, but they are numerous and language descriptions are not consistent in the inclusion and organization of this kind of information. For these reasons, the analysis of markedness criteria in this study is limited to structural coding and inflectional potential.

As it applies to parts of speech, the criterion of inflectional potential must also specify what categories of inflection are relevant. The greatest inflectional potential is found at the prototypes—prototypical combinations of broad semantic class and propositional act function—and different inflectional categories show up in each of the prototypes. Table 5 lists the prototypical inflectional categories for each propositional act function. From these prototypes, inflectional categories can extend vertically (in terms of Croft's conceptual space in Figure 1), applying to the use of the other broad semantic classes in the same propositional act function; or they can extend horizontally, applying to the use of the same broad semantic class in the other propositional act functions. Because the methodology of this study is to compare concepts (represented by corresponding lexemes in particular languages) to each other in identical contexts, only the vertical extensions of inflectional categories described above are considered in this study. The context for comparison

[3] I found only the occasional lack of information about whether or not it is possible.

of concepts is a particular propositional act function in a particular language, so only the inflectional categories prototypical to that propositional act function are relevant. Horizontal extensions of inflectional categories are residual; a lexeme may retain inflectional potential from its morphosyntactic expression in the prototype even when it is used in other propositional act functions. Since these inflectional categories are not equally available to all concepts (because they have different prototypes), they are not appropriate measures for comparing across concepts (see also 4.3.2).

Table 5. Prototypical inflectional categories for each propositional act function (Croft to appear)

Reference	***Modification***	***Predication***
size number definiteness deixis case	degree indexation linker	voice aspect tense modality evidentiality speech act polarity indexation

To summarize, marked values of a category require equal or more structural coding and exhibit equal or less behavioral potential than their unmarked counterparts. In a parts of speech study, what are the categories and values relevant for markedness? The category is the conceptual continuum which was described in section 2.2. Since it is a continuum, in theory its values are infinite. For this study, however, the values are the selected conceptual exemplars, and these are represented in each language by the lexemes that encode those concepts. As described in the previous paragraph, the contexts for comparing these values are the propositional act functions. The markedness asymmetries among the conceptual values are expected to be different for each propositional act function, since a different subset of concepts is prototypical in each propositional act function.

2.3.3. Motivations for typological markedness

Typological markedness patterns are motivated by both token frequency (Greenberg 1966; Bybee 1985; Croft 2003) and the tug-of-war between economy and iconicity (Croft 2003: 101-2).[4] The structure of language represents the structure of the world as humans experience it (iconicity). On the other hand, limitations on time and resources in human communication motivate the minimization of language structure (economy). Frequency provides the asymmetry on which these opposing forces can act. Given a category with two values differing in frequency, economy can be maximized while simultaneously retaining a distinction that is iconic of the distinction in the world: the more frequent value can be zero-coded, while the less frequent value remains overtly-coded. This is precisely the pattern that is found for many categories in many languages (but not the only one of course; in many instances, both or neither of the values are expressed overtly). Inflectional potential can be explained in the same way. Through cross-cutting inflectional capabilities, more frequent values display versatility that is representative of the world. The same cross-cutting inflectional capabilities are less useful for infrequent values, so they are lost or fail to crystallize in the grammar.

[4] The discussion here of the motivations for typological markedness is a cursory one. The reader unfamiliar with this literature is referred to the works cited for a more detailed account.

These ideas are summarized nicely in the oft-quoted mantra, “grammars do best what speakers do most” (Du Bois 1985: 363).

The same principles are at work shaping parts of speech systems despite the countless complexities and intervening factors that exist within such a broad characterization of language structure. Certain concepts—based on their semantic make-up—lend themselves to use in particular discourse functions as a direct reflection of how humans experience the world. Because of their usefulness, these concepts are used more frequently in certain propositional act functions than others. Iconic and economic motivations turn this asymmetry in frequency into an asymmetry in form.

Now that it is sufficiently clear what is meant by typological markedness, I must reiterate that markedness asymmetries are the variation upon which the multidimensional scaling analysis is based. Not every difference in form qualifies as a markedness asymmetry. Consider the following hypothetical example: one lexical class of a language requires a particular derivational monomorphemic suffix in reference, while another class requires a different monomorphemic suffix in reference. The relevant criterion here is structural coding, but these two constructions have the same number of morphemes. In markedness terms, they are equal. There are many linguistic analyses for which any difference in form is significant, but only markedness asymmetries are relevant for this study.

2.4. Sources of the prototype structures

The prototype structures of parts of speech emerge from markedness asymmetries *within* and *across* languages.

The morphosyntactic distinctions made within a particular language serve to divide up lexemes into classes and subclasses. For all languages in this study, morphosyntactic evidence was found to motivate at least a basic three-way distinction for the encoding of objects, properties, and actions. Both structural coding and inflectional potential play a significant role in how these broad semantic classes are realized in each of the three propositional act functions of reference, modification, and predication. Yet, the basic distinctions made in just one language are not enough to fully illustrate prototype structures. Since each lexeme is evaluated for structural coding in three propositional act functions, there is the possibility for some staggering of categorization lines across these environments. However, the distinctions that form lexical classes and subclasses in a particular language are often fairly persistent across propositional act functions. That is, Nouns of a language pattern like other Nouns in reference, and they also pattern like other Nouns in modification. It is less frequent that the class of Nouns that display particular morphosyntactic characteristics in reference is significantly different from another class of Nouns that display particular morphosyntactic characteristics in modification.

However, inflectional potential plays an important role in adding to the markedness structure *within* a particular language. Even as morphosyntactic evidence is found to motivate broad lexical classes within a language (corresponding roughly to objects, properties, and actions) inflectional potential often provides further delineations within each of these classes. More prototypical members of each of these classes can exhibit inflectional capabilities that distinguish them from less prototypical members. Unlike the broad structural coding distinctions that are reinforced in more than one propositional act function, these inflectional potential asymmetries are often specific to a particular propositional act function and its relevant cross-cutting inflectional categories. Thus, even within a single language, major and minor markedness distinctions can be observed. Even still, full prototype structures are not readily apparent until multiple languages are considered.

When data from many languages are compiled, the gradience of the prototype structures emerges. Languages make different decisions on where to draw the lines between both broad lexical classes and smaller subclasses. There is a set of functional pressures that are at work in all languages, and these functional pressures motivate similarities in their lexical classifications. Yet with thousands of lexemes (or even the forty-nine included in this study), no two languages will come to precisely the same lexical classifications. Languages more often align in their classification of some lexemes, and these form the hearts of the prototypes; languages less often align in their classification of other lexemes, and these form the peripheries of the prototypes. It is the accumulation of all these language-specific decisions from many *different* languages that results in the rich prototype structures illustrated in this thesis.

2.5. Cutting up the conceptual space

2.5.1. Semantic maps

An important tool developed in linguistics for representing cross-linguistic regularities in semantic structure is the semantic map model, exemplified in Figure 1. It came to prominence through early work by Anderson (1974, 1982, 1986, 1987), and has since been used by various scholars to represent data from a wide range of linguistic domains (e.g., Croft, Shyldkrot & Kemmer 1987; Kemmer 1993; Stassen 1997; Haspelmath 1997a, 1997b, 2003; van der Auwera & Plungian 1998; Croft 2003). Following Croft (2001: 93), I will use the term 'conceptual space' for the language-universal conceptual structure and 'semantic map' for the language-particular semantic structure that can be 'mapped' onto the conceptual space. The strengths of a conceptual space lie in its ability to capture similarities across many languages without assuming universal language categories and to illustrate those relationships beyond a simple linear structure (Croft 2001; Haspelmath 2003). While some applications of the semantic map model claim only a characterization of the shape of the data, conceptual space—as its name implies—likely reflects a universal conceptual structure in the minds of human beings (Croft 2001: 105, 2003: 138-9).

The relationship between conceptual space and markedness patterns realized in structural coding and behavioral potential has also been explicitly framed (Croft 2003: 140-2). This would suggest that the semantic map model could be applied to the parts of speech data presented here. However, the semantic map model faces some inherent challenges. First, building a conceptual space is not easily tractable with large amounts of data. Second, the semantic map model is unable to deal with exceptions. Current research in typology has increasingly come to appreciate the prevalence and significance of statistical universals, so the ability to deal with 'messy' data is quickly becoming an essential requirement of typological models. Third, the semantic map model is not mathematically formalized. Its nodes and connections fail to represent the *degree* of similarity between related concepts. The existence of a connection between two concepts is captured in a conceptual space, but not the strength of that connection. Fortunately, there is a mathematically-formalized alternative to the semantic map model which has been adapted for and demonstrated on linguistic data.

2.5.2. Multidimensional scaling analysis

Multidimensional scaling (MDS) is multivariate method that directly models similarity in data by distance in a spatial model. It has only recently been adopted from related disciplines for use in grammatical analysis (Croft & Poole 2008; further background on MDS can be found in Poole 2005: ch. 1). MDS models have several advantages over semantic map models. MDS is a mathematically formalized approach that captures not only the categorical relationships among data points, but also their degrees of dissimilarity. In fact, MDS delivers a true Euclidean model. This is an advantage of MDS over alternative methods for discovering structure in multivariate data sets, such as factor analysis and principal components analysis, which only attempt to capture *most* of the variation in the data (for a more detailed comparison of these approaches, see Croft & Poole 2008: 12-13; Baayen 2008: ch. 5).

Not only can MDS be applied to large data sets, the model actually improves with the addition of data (Croft & Poole 2008: 31-2). The data on parts of speech corroborates this claim. As was mentioned above, the inclusion of eleven languages and nearly fifty lexemes for each language turned out to be just enough data to produce recognizable prototype structures and illustrate their internal structures. The coded data for this study comprise a 49 x 169 matrix. By comparison, Haspelmath's (2003) pronoun data (as analyzed by Croft & Poole [2008]) form a 9 x 139 matrix and Dahl's (1985) tense and aspect data form a 250 x 1107 matrix. Finally, MDS can handle exceptions in the data—that is, it measures the fit of the model to the data. I turn now to a brief description of the MDS spatial model that will be important for interpreting the results from the parts of speech data.

An MDS spatial model consists of an arrangement of data points and linear cutting lines. Each cutting line signifies a discrete binary categorization of the data, in this case a single grammatical distinction from a language. The data points represent the conceptual targets used in this study. A single cutting line will divide some portion of the data points from the rest. The points on one side of this line are all the lexemes from a particular language that assume a marked form in a particular propositional act functions, while the points on the other side are all the lexemes that assume the unmarked form. It is the accumulation of all these cutting lines and their discrete categorizations that force the data points into a particular spatial arrangement in the model. Specifically, the cutting lines form polytopes that contain the ideal location for each data point, but a data point can be anywhere within that polytope. More data result in more cutting lines, and more cutting lines result in more accurate polytopes. If two points are close in proximity in the analysis, it is because they fall on the same side of a cutting line more often than not. If two points are very distant, it is because they fall on opposite sides of a cutting line more often than not. In this way, the accumulation of binary categorizations (cutting lines) results in similarity (or dissimilarity) measures among all the data points. For parts of speech, this tool provides a more nuanced illustration of cross-linguistic patterns.

Like the semantic map model, an MDS analysis of linguistic data is hypothesized to capture more than just patterns in the data. A key distinction is made between the language- and construction-specific nature of the cutting lines and the universal nature of the resulting conceptual space. This conceptual space modeled by MDS is hypothesized to be the same for all speakers. According to Croft & Poole, "the relative position and distance of points in the spatial model represent a conceptual organization, presumably the product of human cognition and interaction with the environment, that constrains the structure of grammar" (2008: 33). It can then be inferred that the semantic or functional categories reflected in the structure of this conceptual space are relevant to grammar. In other words, the dimensions of an MDS analysis are meaningful, and they allow for a theoretical interpretation.

The space in the MDS analysis of parts of speech could plausibly represent "conceptual discreteness" (Croft & Poole 2008: 20-1). However, based on the complexity of the data and the theoretically unlimited number of possible lexemes across languages, it is more likely that the space represents a continuum into which more lexemes could fall should they be included in the analysis. Croft & Poole (2008: 21) write that with the accumulation of more data, "what at first appears to be a discrete conceptual space is revealed to be a more continuous structure in a conceptually meaningful spatial representation."

The spatial model is constructed based on boundaries rather than prototypes (Croft & Poole 2008: 11), and this has implications for theoretical interpretation. Prototype structures—centers of similarity gravity around which points cluster—may show up in an MDS analysis, but they are an indirect product of accumulated *dissimilarities*. Therefore, my subsequent discussions of prototype structures in the analysis of parts of speech data should be understood in this light. This fits nicely with the larger conception of language universals as indirect constraints on grammatical variation rather than universal linguistic categories (Croft & Poole 2008: 32). The semantic continuum from object to property to action concepts statistically constrains the possibilities for lexical categorization within a language, even though no two languages choose exactly the same categories and boundaries.

3. Data and methodology

3.1. Selection of the language sample

3.1.1. Genetic and geographic diversity

Figure 2. Geographic distribution of the language sample

For this study, data was collected from eleven genetically and geographically diverse languages. As concerns genetic relatedness, only language families that have been thoroughly demonstrated and accepted by a majority of the linguistic community are considered here. According to this standard, no two languages used in this study belong to the same language family. However, more controversial claims about larger language groupings, such as Amerind (Greenberg 1960, 1987; Ruhlen 1991), have the potential to subsume more than one of these languages. Geographically, three languages were selected from each of three regions—the

Americas, Africa, and Australia/Oceania—and two from Eurasia. Within each region geographic diversity was preferred, with one noteworthy exception: the inclusion of both Mali (Baining) from East New Britain and Toqabaqita from the Solomon Islands. Even though they are genetically unrelated, both are spoken on islands just to the east of New Guinea. Ultimately, the combination of genetic and geographic diversity yields a fairly balanced typological survey.

Table 6. Summary data for the languages used in this study (genetic classification and population numbers are from Ethnologue)

Region	***Language***	***Classification***	***L1 Speakers***
Africa	Kanuri	Nilo-Saharan	3,760,500
	Khwe	Khoisan	6,790
	Kisi	Niger-Congo	10,200
Americas	Creek (Muskogee)	Muskogean	4,000
	K'ichee'	Mayan	2,330,000
	Quechua (Huánuco)	Quechuan	40,000
Australia/Oceania	Jingulu	Australian	52
	Mali (Baining)	East New Britain	2,200
	Toqabaqita	Austronesian	12,600
Eurasia	English	Indo-European	335,148,868
	Ingush	North Caucasian	322,900

Admittedly, the availability of detailed reference materials was a primary factor in the selection of these languages. For most of the languages in this study, descriptive grammars and two-way dictionaries with English were required to collect the necessary grammatical and lexical data. Many of these were recommendations from experienced typologists, and a few languages originally included were replaced due to the insufficient nature of their reference materials.

Typological diversity was considered in the selection of languages, but not all typological information was expected to be relevant. For example, there is no expected correlation between word order and parts of speech strategies.

3.1.2. Diversity in parts of speech strategies

Another form of diversity that is important for this particular study is that of parts of speech strategies. By this, I am referring to both the claims made of some languages that they do not have all three basic parts of speech, as well the generalizations made about other languages based on perceived similarities between two of the three major parts of speech. Languages "without adjectives" are the most common example of the first type. Additionally, there are languages for which it has been suggested that the Noun/Verb distinction does not exist. The second type includes languages with "Verbal Adjectives" or "Adjectival Nouns" (see section 2.2.2). The generalization here is that the lexical class of Adjectives in a particular language shares a significant amount of morphosyntactic properties with either Nouns or Verbs, in some cases to the extent that they might be considered one large lexical class with two subclasses. Both types are represented in this study.

Another cross-linguistic variable is the size of the Adjective class (also discussed in 2.2.2). Again, the languages used in this study cover the range of inventory sizes, from a language with one 'true' Adjective to others with very large Adjective classes.

It would be difficult to group the languages of my study definitively into the types described above. While several of the grammar authors for these languages make such claims, the strengths of the claims vary, as does the amount of evidence that is provided by the authors to support their claims. To illustrate the diversity of the language sample in terms of parts of speech strategies, I have included relevant excerpts from the language resources in Table 7. However, it cannot be understated that markedness asymmetries of at least one type were found to distinguish all three semantic classes (objects, properties, and actions) in every language in the study. That is, there was no language for which I failed to find morphosyntactic evidence to distinguish at least three basic parts of speech from each other.

Table 7. Selected insights into the nature of parts of speech from the authors of the reference materials

Language	***Parts of Speech Description***
Kanuri	"Words that translate into other languages as adjectives… have no overt phonological, morphological or syntactic characteristics that differ from the same characteristics of words that translate into other languages as nouns." (Hutchison 1981: 35)
Khwe	"Khwe does not have a coherent, proper word class of adjectives… It formally distinguishes between verbal adjectives, adjectives derived from nouns, adjectives related to pronouns, and numerals" (Kilian-Hatz 2008: 195)
Kisi	"Adjectives are an easily defined class consisting of words attributive in meaning and showing agreement with nouns they modify." (Childs 1995: 151)
Creek (Muskogee)	"Stems in Creek can be divided into several classes based on their grammatical behavior" (Martin 2011: 29)
Quechua (Huánuco)	"There is insufficient evidence of a strictly morpho-syntactic nature for considering that nouns and adjectives form distinct lexical classes." (Weber 1989: 9)
Jingulu	"Australian languages often do not show any morphosyntactic contrast between nouns and adjectives… In Jingulu, however, there appear to be some reasons for considering nouns and adjectives to be at least partially distinct parts of speech." (Pensalfini 2003: 57)
Mali (Baining)	"It can be difficult to distinguish adjectives from verbs since an adjective may head an intransitive predicate." (Stebbins 2011: 95)
Toqabaqita	The class of adjectives consists of one member (Lichtenberk 2008b: 52, 230)
Ingush	Ingush has three major classes of words: nouns, verbs and modifiers (Nichols 2011: 113)

3.2. Data sources

Three different methods were used in the collection of data from eleven languages. For nine of the eleven languages, data was collected from grammar and dictionary references. As mentioned above, the quality of reference materials was important in the selection of languages for the study. Even so, challenges arose with this method of data collection, and they are discussed later in this paper. As a native speaker, self-elicitation and personal grammaticality judgments were used for English.

The K'ichee' Mayan data in this study were collected uniquely through elicitation, so I describe the methods here. Much of the elicitation was from James Mondloch, Professor of K'ichee' at the University of New Mexico. For certain constructions, he deferred to his wife, Maria Mondloch, who is a native speaker of the language.

The methods of elicitation evolved throughout the collection of the data. Before any elicitation, a brief overview of this study was presented to Dr. Mondloch. It was explained that the analysis was of particular lexemes and how these lexemes are encoded in each of the targeted propositional act functions. Originally, example sentences were presented in English with the lexeme appearing in all capitals in its unmarked form:

<<WHITE represents purity>> (reference)
<<The WHITE dog barked>> (modification)
<<The cat WHITE>> (predication)

The goal of this elicitation strategy was to avoid a bias toward English constructions, particularly for the non-prototypical pairings of semantic class and propositional act function. For example, if a gerund form was used in the elicitation prompt for an action in reference, it might bias the speaker to use an analogous form in K'ichee' at the expense of other possible morphosyntactic strategies for this combination.

Ultimately, this method was insufficient. For many combinations of lexeme and non-prototypical propositional act function, the bare (un-derived) form of the lexeme created confusion as to the meaning of the sentence. To clarify the intended elicitation, the lexemes had to be given in fully grammatical (derived) form in the English prompts. For English prompts for which there is more than one possible morphosyntactic strategy, all relevant strategies were given in the elicitation. For example, English has multiple strategies for encoding an action word in modification, namely a Participle form and a Relative Clause construction. Both of these were given in English to alleviate bias in eliciting the same combination in K'ichee':

<<The FIGHTING man shouted>> (Participle)
<<The man WHO FIGHTS/WHO WAS FIGHTING shouted>> (Relative Clause)

This revised elicitation strategy was much more effective and was responsible for most of the K'ichee' data.

The difficulty of the elicitation process for K'ichee' actually supports the theory of parts of speech to which this paper ascribes. On the surface, frequency can explain why certain combinations of semantic class and propositional act function are difficult to artificially construct for the respondent. The more a particular construction is practiced, the more effortless and natural it becomes to produce. As was discussed in section 2.3.3, some combinations of semantic class and propositional act function are very uncommon. Furthermore, the elicitation of a *particular* lexeme in one of these uncommon combinations may be asking the respondent to produce a form of that lexeme never produced before. For example, it's very possible that a native speaker of K'ichee' has never heard or produced the abstract referential derived form of ROUND equivalent to the (rather awkward) English *round-ness*. The respondent will likely be able to produce such a form based on analogy with other lexemes, but hesitation and/or doubt as to the grammaticality of the new construction would be understandable. In fact, this reaction was well attested in my elicitation of the K'ichee' data. For many of the non-prototypical combinations, the respondents made comments to the effect of, 'I don't know if I've ever heard that before.'

3.3. Attested patterns of markedness

The purpose of this section is to present examples of asymmetry in markedness from the languages in this study. These asymmetries—captured in the data coded for multidimensional scaling—are what differentiate the concepts in the spatial model. The asymmetries are realized in both structural coding and inflectional potential. They will be presented for each propositional act function in turn. Not all the asymmetries in markedness that were found in the data are presented here. Rather, this section features a subset of the relevant data to illustrate the observations of markedness that were found among the eleven languages of this study.[5]

3.3.1. Reference

An example of structural coding in reference comes from Ingush. Abstract Nouns can be derived from Adjectives with the suffix –*al*. Several examples of this derivation are illustrated in (1).

---

[5] Many examples are included in this paper to illustrate its findings. All include an interlinear morpheme translation. For all these languages, examples are reproduced here using the same phonetic, phonemic, or orthographic script used in source material. No attempt was made to convert the data to a standard transcription such as the International Phonetic Alphabet. To my knowledge, this decision bears no impact on the morphological and syntactic variables that are at issue.

(1) Abstract Noun derivation in Ingush (Nichols 2011: 158)

Adjective:	Abstract Noun:
dika 'good'	*dikal* 'goodness'
d.aza 'heavy'	*dazal* 'heaviness, weight, value'
q'eana 'elderly'	*q'eanal* 'old age'
q'uona 'young'	*q'uonal* 'youth'

This derivation even works for the word *eghazal* 'anger' from the verb *eghaz-lu* 'get angry', even though there is no corresponding adjective form for 'angry'. Therefore, with regard to structural coding, members of the Ingush Adjective class are marked in reference relative to members of the Noun class.

The inflection of number on Subject Markers in Toqabaqita illustrates the effect of animacy on markedness of objects in reference. In the third person, higher animate nouns are almost always individuated, and dual and plural Subject Markers are used. On the other hand, a singular Subject Marker is often used for even dual and plural mentions of lesser animates and inanimates. There are exceptions to this rule in both directions.

(2) Animacy distinctions in Subject Marker number agreement in Toqabaqita (Lichtenberk 2008: 150)

a.

botho	*baa*	*ki*	*kera*	*oli*	*na=mai*
pig	that	PL	3PL.NFUT	return	PRF=VENT

'The pigs have come back.'

b.

fau	*neqe*	*ki*	*qe*	*kuluqa*
stone	this	PL	3SG.NFUT	be.heavy

'These stones are heavy.'

Furthermore, pronouns are not often used when the referent is not human (Lichtenberk 2008: 72). These distinctions in inflectional potential serve to disambiguate some object concepts from others in Toqabaqita.

3.3.2. Modification

In this section I present some examples of markedness distinctions in modification. A straightforward example comes from Kanuri, in which Nouns are overtly marked with a suffix when used in modification. For example, the *-má* suffix in (3) indicates a relationship of trade, profession, or origin, among others.

(3) Kanuri Nouns in modification (Hutchison 1981: 196-7)

kâm *fər-má*
person horse-DENOM
'person owning a horse'

Modification constructions in Toqabaqita demonstrate how labels such as Noun, Adjective, and Verb do not tell the full story of markedness distinctions. According to Lichtenberk (2008), Toqabaqita has only one Adjective. This analysis is based on inflectional potential: this sole Adjective is the only lexeme that has different forms based on number, animacy, and count/mass noun status. As a result, most property words in Toqabaqita are categorized as Verbs. Yet, the semantic distinction between properties and actions within this class of Verbs is still realized in structural coding. Only Verbs with "a stative meaning in their semantic range" (327) can be used as modifiers without any additional structural coding.

(4) Toqabaqita Stative Verbs in modification (Lichtenberk 2008: 328)

toqa *maamaelia* *toqa* *suukwaqi* *toqa* *leqa*
people be.powerful people be.strong people be.good
'powerful people, strong people, good people'

All other Toqabaqita Verbs can only modify referents in a relative clause introduced by the particle *na*. The placement of the relative clause also serves to differentiate this type of modification from the stative Verb modifiers.

Even more, Toqabaqita Nouns can modify other Nouns in a genitive relationship with a possessive suffix or simple juxtaposition (Lichtenberk 2008b: ch. 8).

(4) Toqabaqita genitive construction (Lichtenberk 2008b: 385)

kaleko	*faalu*	*wela*
clothes	be.new	child

'the child's new clothes'

In summary, it is possible for all Toqabaqita object and property words to assume a modification role without any added structural coding. Only the non-stative action words require additional structural coding to do the same. Despite the classification of almost all property words as Verbs in Toqabaqita, a markedness distinction remains between property-like (stative) Verbs and action-like (non-stative) Verbs. The same large Verb class gives a "verby" impression of the property words in Toqabaqita, yet it is object words that pattern with the property words as unmarked in modification with regard to their structural coding.

Having demonstrated how markedness differences manifest in structural coding, I turn now to inflectional potential in modification. The Adjective class in Ingush makes distinctions within the broad semantic class of property words based on their inflectional potential in modification. Only two property words—'big' and 'small'—agree in number with the Ingush Noun. Furthermore, less than ten percent of Adjectives agree in gender with the Noun.

(5) Number and gender agreement in Ingush (Nichols 2011: 219-21)

a.

v-oaqqa	*sag*
V-big	man

'old man'

b. *b-oaqq-ii* *nax*
B-big-PL people
'elders'

The inflectional capabilities of these Adjectives set them apart from other members of the Ingush Adjective class.

3.3.3. Predication

Structural coding in predication is typically manifested in the form of a copula for objects and/or properties. For some languages, such as Kanuri, no copula is necessary. Actions, objects and properties can be predicated by juxtaposition to the subject.

(6) Predication of objects and properties in Kanuri (Hutchison 1981: 10-1)

a. *Álì* *bàrèmà*
Ali farmer
'Ali is a farmer.'

b. *Fə̂r-nzé* *kúrà.*
horse-3SG.POSS big
'His horse is big.'

For other languages, the copula is required. In Mali, the coordinator *da* 'and' functions as a copula for the predication of properties. This construction distinguishes properties as more marked than actions in predication.

(7) Predication of properties in Mali (Stebbins 2011: 49-50)

chēvicha *da* *aululka*
kēvi=ka *da* *aulul=ka*
CONTR=M.SG.III and tall=3SG.M.III
'That guy is tall.'

A unique pattern of markedness is found in Jingulu predication and it implicates structural coding. The Jingulu predicate consists of a mandatory Light

Verb and an optional Coverbal Root. The least marked lexemes in predication are the Light Verbs themselves, since they occur with minimal structural coding. The Coverbal Roots are more marked because they only occur with an additional morpheme—a Light Verb.

(8) Verbal elements of Jingulu (Pensalfini 2003: 59-60)

a. *ya-jiyimi* *bininja*
3SG-come man
'The man is coming.'

b. *ngaba-nga-ju* *karnarinymi*
have-1SG-do spear
'I have a spear.'

c. *ngaba-nga-rriyi* *karnarinymi*
have-1SG-will.go spear
'I'll take a spear.'

The unique way in which Mali Verbs inflect for tense in predication constructions results in several levels of markedness. The inflectional capability of a Mali Verb depends on both its class (A-D) and whether it is "Stative Intransitive" or "Active Intransitive"/"Transitive" (Stebbins 2011: 53-7). The Verb classes are differentiated by their phonological forms and the ability of the verb to inflect for tense directly: Class A Verbs encode three tenses, Class B and C Verbs encode only the present/non-present distinction, and Class D Verbs do not encode any tense distinctions. The stativity and transitivity categories are relevant because they determine which Concordial Pronoun accompanies the Verb; Stative Intransitives require a Class III Concordial Pronoun which does not encode a past/non-past distinction like other Concordial Pronoun classes.

Since both the ability to inflect and the nature of that inflection (ablaut, affixation, or periphrasis) are relevant factors in markedness, Mali Verbs feature five levels of markedness.

(9) Levels of markedness in tense inflection of Mali Verbs, from least marked to most marked (Stebbins 2011)

a. Class A: Encode past/present/future on Verb

b. Class B+C (Transitive/Active Intransitive): Encode present/non-present on Verb, and past/non-past on Concordial Pronoun

c. Class B+C (Stative Intransitive): Encode present/non-present on Verb, but no past/non-past distinction due to use of Class III Concordial Pronoun

d. Class D (Transitive/Active Intransitive): No present/non-present distinction on Verb, but encode past/non-past on Concordial Pronoun

e. Class D (Stative Intransitive): No present/non-present distinction on Verb, and no past/non-past distinction due to use of Class III Concordial Pronoun

The third level of markedness featuring Classes B and C Stative Intransitive Verbs is not represented by any of the lexemes in this study, but the other four levels are attested. Table 8 maps these levels onto the intersection of Verb Class and transitivity/stativity.

Table 8. Tense distinctions of Mali Verbs according to Verb Class, transitivity, and stativity (based on data and description in Stebbins 2011)

<table>
<tr><th></th><th>Transitive</th><th>Active Intransitive</th><th>Stative Intransitive</th></tr>
<tr><td>Class A</td><td colspan="3">PST/PRS/FUT</td></tr>
<tr><td>Class B</td><td colspan="2" rowspan="2">PST/PRS/FUT (partially periphrastic)</td><td rowspan="2">PRS/NPRS</td></tr>
<tr><td>Class C</td></tr>
<tr><td>Class D</td><td colspan="2">PST/NPST</td><td>No Tense Distinctions</td></tr>
</table>

Inflectional potential also yields markedness distinctions in predication in Creek. Aspect and other distinctions are made on the Creek Verb via root alternations ("grades") and suffixes. Creek has a primary distinction between temporary events (Non-Duratives) and states/prolonged events (Duratives). All unmarked Verbs (includes Adjectives, considered a subclass of Verbs) in Creek refer to an event, while Verbs with a Durative suffix can be either eventive or stative (Martin 2011: 248-51). Many of the action concepts in this study are specifically non-eventive (e.g., KNOW, WEAR) and would require this Durative suffix—marked structural coding—in order to accomplish a non-eventive meaning in predication. This would indicate a markedness distinction between eventive and non-eventive concepts in Creek.

However, there is another available construction for concepts that fall somewhere between non-eventive and prototypically stative: "Verbs referring to dressing, knowledge, perception, and holding… commonly occur in the resultative stative aspect" (244). The Resultative Stative is close in meaning to the Durative, but there are subtle differences; the Resultative Stative is used for events that are portrayed as shorter in duration, while the Durative is used for longer events or those that are more neutral in this regard (247). The form of the Resultative Stative features a falling tone, but no suffix like the Durative form. Therefore, these lexemes are unmarked in predication just like the eventive concepts.

In Table 9, the cells with bold text are the aspectual forms that match the conceptual targets of this study. Both 'know' and 'wear' represent concepts included in my study. While they are not concepts included in my study, 'ripe' and 'bark' are used here to represent physical properties and cyclic achievements, respectively;

this was a matter of convenience, as the Creek forms for these lexemes were readily available.

Table 9. Aspectual distinctions in Creek Verbs (from data and description in Martin 2011)

	Non-Durative Eventive (Lengthened Grade)	***Resultative Stative*** (Falling Grade)	***Durative*** (Zero Grade + Durative Suffix)
'ripe'	*'lo:kc-ís* 'it's getting ripe'		***lókc-i:-s*** **'it's ripe'**
'know'	*ki:ll-ís* 's/he is learning it'	***kîll-is*** **'s/he knows'**	*kíll-i:-s* 's/he knows' or 's/he is knowledgeable'
'wear'	*'a:cc-ís* 's/he is putting on'	***â:cc-is*** **'s/he is wearing'**	*ácc-i:-s* 's/he has on' or 's/he would/could wear'
'bark'	***'wo:hk-ís*** **'it's barking'**		*wo:hk-í:-s* 'it barks (all the time)'

3.4. Coding the data

The multidimensional scaling analysis used in this study was accomplished with the MDS OC R program, developed by Keith Poole and modified slightly by Jason Timms for use on linguistic data. This program features Poole's Optimal Classification algorithm, a nonparametric binary unfolding algorithm which progressively approximates data points and cutting lines until an optimal classification is reached (Poole 2000, 2005).

The MDS analysis is constructed from binary data, and each column of data is responsible for one cutting line in the analysis. Each column represents one level of markedness attested for at least one lexeme in a particular language and propositional act function. The forty-nine conceptual targets (represented by lexemes in each language) constitute the rows, and they are coded in that column for one of two values indicating whether or not they occur at that level of markedness in that language and propositional act function.

I will illustrate this with an example. One column of my data is labeled with the following shorthand: “KI-R-SC-Affix”. The shorthand contains four components separated by dashes. The first component is a two-letter code indicating the source language, in this case KI for Kisi. The second component indicates the propositional act function in question (R=reference, M=modification, P=predication). The third component indicates whether the markedness distinction in question pertains to structural coding (SC) or behavioral potential (BP). The fourth component can indicate the type of structural coding, or—for the criterion of inflectional potential—the category being inflected for. For the example above, “Affix” is the amount of structural coding necessary for the lexeme to occur in that propositional act function. There are a few column labels that include a fifth component, necessitated by differences in *how* inflectional potential is accomplished, either via internal change, affixation, or periphrasis (see 2.3.2).

A third coding option can indicate when the data is not available. This may result from any of a number of scenarios. For some languages, a particular lexeme could not be found to match the conceptual target. That meant the lexeme was assigned this value across all the columns for that language. In other cases, the grammar writer was explicit that s/he had no information regarding a combination of lexeme(s) and propositional act function(s). Finally, there were some relevant constructions that were left unclear or unmentioned by the reference material.

Fortunately, the MDS algorithm used here can adeptly handle these missing data points.

In many instances, various inflectional abilities pattern together. For example, a set of action words may take on several inflectional abilities in predication, such as tense, aspect, mood, etc. If all these inflectional abilities apply to the same set of lexemes, they are treated together in one column of code as a level of markedness (which at least some other lexemes do not display). This was a step to avoid redundancy; representing each inflection in its own column would have resulted in many identical cutting lines.

3.5. Dimensionality

Once the data is coded and the MDS analysis complete, there is still another theoretical decision to make. The appropriate number of dimensions for an MDS analysis is dependent on the properties of the data. The analysis is performed in one, two, and three dimensions, and fitness statistics are provided to indicate the level of fit for each number of dimensions. Complex data result in an inverse relationship between informativeness and goodness of fit: as more dimensions are included, goodness of fit increases but the informativeness of the model decreases. The best model is the lowest-dimensional model that offers a relatively high degree of fit, and for which each additional dimension offers only small improvements in the fit (Borg and Groenen 1997: ch. 3 & 4, cited in Croft & Poole 2008: 12). For data in which language universals are present, a low-dimensional model with relatively good degree of fit is expected. The limited number of dimensions can then be associated with a small set of functional factors that are responsible for the variation.

4. Methodological challenges

Before a presentation and discussion of the results of this study, it is important to bring attention to some of its methodological and theoretical challenges. While these challenges do not take away from the significance of the results, they appropriately qualify them. For future studies in this vein, this section offers potential areas for critical thought and creative solutions.

4.1. Cross-linguistic comparison

A persistent challenge in the discipline-wide typological endeavor is that of cross-linguistic comparison. In fact, the theory of lexical word classes maintained in this study rejects the universality of formal word classes. As we have seen, a word class such as Adjective in a particular language is not equivalent or comparable to the Adjective word class in another language. Yet this study attempts to relate the formal word class distinctions made in each language—including various constructions and grammatical categories—to the distinctions made in the other languages of the study. The solution is to compare lexemes only to other lexemes in the *same language* and the *same propositional act function*. A lexeme in a particular language is never compared to a lexeme with the same meaning in another language. Only the lexemes within a language are compared to each other, and measures of dissimilarity are summed cumulatively across constructions, propositional act functions, and languages. So overall dissimilarity measures among concepts are the cumulative dissimilarities of many intra-language and within-propositional act function comparisons.

4.2. Scope

The primary logistical challenge of this study is its combination of ambitious scope and attention to detail. Collection of data involved combing through thousands of pages of grammars and dictionaries to find the relevant information for nearly fifty lexemes, in three propositional act functions (and for typically more than one possible construction in each of these combinations), and for eleven languages. Yet all these were necessary, and the interesting results of this study would likely be compromised if not for their inclusion. The approximately fifteen concepts for each broad semantic class are just enough to explore the internal structure of each prototype. All three broad semantic classes had to be included, since it is the behavior of concepts between the prototypes—relative to those at or near the prototypes—that represent arguably the most interesting data points.

4.3. Theoretical constraints

4.3.1. Distributional potential

As discussed in section 2.3.2, distributional potential is not coded directly in this study. However, it influences other markedness patterns that are included.

4.3.2. Non-prototypical inflectional potential

It is important to remind the reader what is *not* captured in this study. In section 2.3.2, I indicated that inflectional potential can extend from the prototypes to the other broad semantic classes or to other propositional act functions. The first of these is included in the scope of this study. The second—inflectional potential that is not prototypical for that propositional act function—is not included.

For example, tense inflection is prototypical for predication, and Creek is no exception: it features a full set of tense inflections in main clauses (actions in predication). Relative clauses (actions in modification) have the same tense distinctions as main clauses (Martin 2011: 398), but complement clauses (actions in reference) have fewer tense distinctions (390). A similar phenomenon can be found in Toqabaqita. When used in modification, Toqabaqita Verbs can retain some inflectional capabilities prototypical for predication; yet there are also some verbal particles that they cannot take in this construction. In (10), the Verb *mae* 'be dead' is modifying 'man' and is inflected for perfective aspect.

(10) Toqabaqita (Lichtenberk 2008b: 331)

kafa	*qeri,*	*wane*	*mae*	*naqa*	*n=e*
comb	this	man	be.dead	PRF	FOC=3SG.NFUT
alu-lu-a					
own-RDP-3SG.OBJ					

'This comb, it was the dead man (lit.: the now-dead man) who used to own it.'

These patterns match what is expected by the prototype theory of parts of speech; as the combination of lexical-semantic class and propositional act function move further from prototypicality, markedness will stay the same or increase (in this case, lost inflectional potential). However, these distinctions are not captured in this study because tense and aspect are not prototypical inflectional categories for modification or reference.

4.4. Practical constraints

4.4.1. Availability of data

Even with excellent grammars and extensive dictionaries at my disposal, the availability of data was a challenge throughout the study. A few examples will suffice to illustrate these challenges.[6] In Kisi, for example, the grammar indicates that the "o" noun class is unmarked on the noun (no suffixed noun class pronoun) (Childs:1988: 231), yet the dictionary for Kisi does not list the noun class for each noun entry. The Mali grammar, despite its clarity and organization, does not give any information about how Adjectives or Verbs are used in reference. Some grammars were more explicit in certain domains; the Kisi reference grammar left out discussion of the genitive due to its intended focus on the phonology and morphology of the language.

4.4.2. Generalizability and detail

The nature of grammars and dictionaries additionally presents the problem of categorization and generalization. For this study, it is often the exceptions to the rules that represent the most interesting data points. Yet grammars and dictionaries must necessarily generalize in order to capture as many important facts about a language as possible within reasonable constraints of time and space set by themselves, publishers, and practicality. I found occasional examples in my research

[6] These examples are in no way intended to criticize the authors of the source materials. On the contrary, the typological nature of this research is only possible due to the diligence of fieldworkers and descriptive linguists. The grammars and dictionaries utilized were chosen for their high quality.

that indicate an inconsistency among lexemes assigned to the same class or category. An interesting example comes from K'ichee'. The suffixes –*iil* and –*aal* often make Abstract Nouns out of Adjectives, such as *utz-iil* 'goodness' from *utz* 'good', *nim-aal* 'bigness' from *niim* 'big', and *k'a'n-aal* 'angriness/anger' from *k'a'n* 'angry'. However, *ketekik* 'round' yields *ketekik-iil* 'spherical'—an Adjective just like its stem but with an added spatial dimension in its meaning. In this case, generalizations regarding the suffix fail to capture what constitutes a significant distinction in my analysis.

In fact, the languages that are most interesting for this kind of study are those that make various distinctions among and within major lexical classes. Those same languages, and in particular their various gradations, are the most difficult for the descriptive linguist to accurately describe without extremely detailed attention to the unique properties of individual words and constructions.

4.4.3. Gradience in grammaticality

Not surprisingly, the elicitation of K'ichee' Mayan data brought some unique insights. The most important of these is the gradient nature of grammaticality. As I have argued above, the broad semantic classes are not distinguished categorically, but the lexemes of these classes fall on a continuum based on semantic primitives. This continuum was evident in the elicitation responses for K'ichee'.

The semantic continuum was perhaps best exemplified in the K'ichee'an strategies for predicating property and object words. The responses and subsequent grammaticality judgments included a range of certainty with more than a couple "awkward" but grammatically acceptable responses. Much of the uncertainty revolved around the use of *aree*, which—in addition to its other functions—serves essentially as a copula in these examples. Generally, the copula is necessary for predicating Nouns, but not for predicating Adjectives. However, the responses to elicitation indicate that the distinction is not as categorical as it seems. (11) shows the predication of three lexemes in K'ichee'.

(11) Object and property predication in K'ichee' Mayan (Mondloch, pers. comm.)

a. <<The woman is GOOD>>
utz le ixoq

utz	*le*	*ixoq*
good	DEM	woman

'The woman is good.'

b. <<The alligator's home is RIVER>>
(aree) ja' ro'ch le ayiin

(aree)	*ja'*	*r-o'ch*	*le*	*ayiin*
(COP)	river	3.POSS-home	DEM	alligator

'The alligator's home is the river.'

c. <<The lazy one is WOMAN>>
aree ixoq le q'or

aree	*ixoq*	*le*	*q'or*
COP	woman	DEM	lazy(one)

'The lazy one is the woman.'

The examples above are ordered such that the lexeme being predicated goes from most prototypically property-like to most prototypically object-like. While (11a) and (11c) fit the generalization for copula use given above, the copula *aree* is optional in (11b).

For the most prototypical property words, such as example (11a) the copula is not required. These elicitations presented no problem for the K'ichee' speaker. However, for lexemes that fall somewhat intermediate to the property and object prototypes, it became more difficult to determine whether the copula was necessary or not. Moving closer to the object prototype, 'river' in (11b) elicited a grammatical sentence both with and without the copula; however, this sentence without the copula was approaching ungrammaticality. Finally, (11c) shows that the most prototypical object words require the copula in predication—it is ungrammatical to leave it out. Moreover, this sentence was still considered to be somewhat awkward even when the copula was included.

This type of gradience in grammaticality is likely present in some form in all the languages used in this study. I found only a couple mentions of it in the resources, however. As discussed above, the authors of language resources are forced to generalize and categorize for the sake of economy and readability. As a result, much of this kind of variation is not represented in this study.

4.4.4. Variation that is not captured in the study due to list of concepts

In some cases, the list of lexical-semantic concepts that I chose for this study resulted in the omission of linguistic variation. A clear example of this comes from Toqabaqita's lonely Adjective class. There is only one Adjective in Toqabaqita—'small'—and it is not a concept included in this study. This word comes before a Noun and it has three forms that vary according to its number, animacy, and count/mass status.

(12) Toqabaqita's only Adjective (Lichtenberk 2008: 230)

kasi	*biqu*	*faqekwa*
small	house	be.small

'(very) small house'

Unfortunately, the inflectional potential that makes this lexeme unique in Toqabaqita is not represented in my analysis.

Toqabaqita also features two basic classes of transitive Verbs (Lichtenberk 2008: 51). In one class, only third-person direct objects are indexed on the Verb. In the second class, all direct objects are indexed. The lexemes used in this study include only transitive Verbs of the first class, so this distinction in markedness is not captured in the data either.

4.4.5. Precision in meaning

Another challenge in the collection of data is the proliferation of senses that a word can take on. This was especially apparent among action words in many languages. My predictions for meaningful conceptual differences among lexemes included fine aspectual distinctions, and these are often transcended by the senses of a single word. A common example is the conceptual target WEAR, which is supposed to represent an inactive action of possession. However, in many languages, this concept is represented by a word that also means 'put on', which is surely not inactive.

This polysemy is not in itself a problem for the methodology. After all, if the word can be used for the inactive meaning, then it must be included in the study. However, with English used as a meta-language, the precise meaning of the word in the source language can get lost in translation. English does not make all possible semantic and grammatical distinctions, so English translations do not always convey these subtle distinctions in the source language. If a word is translated to English as ‘sit’, it might refer to the inchoative action of sitting down or the state of being in a sitting position. To be clear, this is not a criticism of the source materials; it would be unrealistic to expect this level of detail in a dictionary that was created with so many other needs and priorities in mind. This reality just means that some assumptions must be made along the way regarding the exact meanings of certain words. One possible solution is to elicit data using images or videos (e.g., Levinson & Meira 2003), but that was not realistic for the scope and timeframe of this project.

4.5. Linguistic considerations

4.5.1. Morpheme counting

A methodological challenge that can be taken for granted in analyzing differences in structural coding is morpheme counting. It can be difficult to determine whether an additional morpheme is present in some cases. A simple example from Kisi is sufficient to illustrate this point. Consider the derivational forms of (13).

(13) Lexical derivation in Kisi (Childs 2000)

Word:	Lexical Class:
lùsùlló ‘be heavy’	Verb
lùsùlló ‘weight, heaviness’	Noun
lùsùl-ɛ́í ‘heavy’	Adjective

There is no difference between the Verbal and Nominal forms, which feature a single *ó* vowel at the end of the word. The Adjectival form, however, replaces this *ó* with a two-vowel *-ɛ́í* suffix. For this study, I accept the morphological analyses provided by the grammar authors, since they are likely informed by their expertise in the linguistic analysis of the language.

4.5.2. Choosing among lexical candidates

Another common challenge is the existence of multiple words in a language for one concept. I was able to narrow down the possibilities in most cases by favoring words with the highest frequency (often one of the options is marked as rare or archaic) and the most general meaning for the target concept. In some cases, more than one lexeme for a concept was retained. If these lexemes exhibited different markedness patterns, that diversity was included in the analysis in the same way as the diversity of a single lexeme that can be used in multiple constructions. For example, Kisi has both an Adjective *kɛ̀ndɛ̀* meaning 'good' and a Verb *nàŋɔ̀ɔ́* meaning 'be good' (Childs 2000). Since these lexemes fall into different classes, they exhibit different markedness patterns, and this diversity has consequences for the placement of this concept in the multidimensional scaling analysis.

4.5.3. Compounds

Compounding and lexicalization present a challenge to any synchronic analysis, including this one. This issue is too broad to include a full discussion here, but the challenge can be illustrated with examples from the data.

In his grammar of Creek, Jack Martin identifies three distinct constructions on the lexicalization pathway from the participle modifying a noun to the fully lexicalized compound. He writes that "words describing color, shape, age, or size commonly combine with the nouns they modify in Creek, as in *ma ifa-lást-i* 'that black dog', but these are distinct grammatically from compounds" (2011: 114). The three constructions are shown in (14a-c).

(14) Noun-Participle Lexicalization in Creek (Martin 2011: 125)

a. Noun + Participle:
ifá *lást-i:*
dog black-DUR
'black dog'

b. Noun + Reduced Participle (adjoined):
ifa-lást-i
dog-black-(DUR)
'black dog'

c. Noun + Reduced Participle (compounded):
fos-cá:t-i
bird-red-(DUR)
'cardinal'

The key to this analysis is semantic: a compound typically has a fixed reference to a type. The most lexicalized of these three constructions in Creek exhibits the fixed reference—in this example 'cardinal' rather than any red bird.

4.5.4. Diachrony

Diachrony also presents a challenge to the kind of synchronic analysis sought in this study in much the same way as the other gradient phenomena described above. With gradual processes of lexicalization and grammaticalization, the question arises as to which historical components of a word are still recognizable to the speaker. In American English, for example, the word *happy* derives historically from the Latin *hap* ('chance, fortune') plus the common adjectivizing suffix –*y*. However, a contemporary analysis likely would not treat *happy* as a multi-morphemic, and most English speakers would not recognize this internal structure either. The question of what is cognitively real to speakers is not always easy to determine, particularly from a grammar or dictionary.

Kanuri's two verb classes are a primary example of this diachronic challenge. Class 2 Verbs include a root *n* which comes from a lexeme meaning 'say, think', and the semantic components of these Class 2 Verbs derive from non-Verbal lexemes. Class 1 Verbs do not include this root.

(15) Verbal morphology in Kanuri (Hutchison 1981: 90-1)

a. Class 1:
búkìn
bù-k-ìn
eat-1SG-IPFV
'I eat'

b. Class 2:
lěngîn
lè-n-k-ìn
go-say/think-1SG-IPFV
'I go'

It is not clear whether the modern speakers of Kanuri recognize the root *n* as a distinct morpheme or as part of a larger root with the semantic component. The analysis of markedness depends on this detail, since Class 2 Verbs have extra structural coding if this root is analyzed as separate. For cases such as these, I defer to the morphological analysis of the grammar. If the grammar presents these components as separate morphemes, then I treat them as such. Although grammar writers may be biased toward an at least partially etymological analysis of morphology because it offers a deeper level of explanation, I am not in a position to determine whether this is actually the case and for which analyses.

4.5.5. Constructional variation

Languages often present multiple solutions to the same problem. This is true also of constructional possibilities for a given lexeme in a particular propositional act function—there is all too often more than one construction available. These possible constructions may be equally available to all lexemes of a certain lexical class or restricted to a certain subset. In these cases, the constructional possibilities afforded a particular lexeme are relevant for this study, especially when one construction is more or less marked than the other. They are included when the information about their applicability is described in sufficient detail. The method for coding data in this study allows a particular lexeme to be analyzed as having more than one level of markedness. In fact, these cases can result in a richer analysis. Each lexeme/propositional act function combination with more than one constructional possibility—one more marked and one less marked—exhibits both similarity and dissimilarity with a lexeme that is *always* less marked, and exhibits both similarity and dissimilarity with a lexeme that is *always* more marked. Thus just two levels of markedness can yield three different lexical classes when one lexical class can be used in both constructions.

In some cases, more than one construction is possible but there are strong correlations between the construction and another predictive factor. An example of this was presented in (2) for Toqabaqita. Higher-animate subjects usually trigger number agreement in the Subject Marker, while inanimate and lower-animate subjects are almost always used with the singular Subject Marker regardless of their number (Lichtenberk 2008: 51). This is just a generalization, and there are exceptions for both groups. Again, for cases like these, I defer to the analysis provided in the grammar. That is, if the generalization is strong enough that the grammar writer is able to identify a boundary, then I include that distinction in my analysis.

Some constructional alternatives are based on other factors, such as the particular relationship of that lexeme to the construction or another element therein. For example, Toqabaqita features several possible constructions for object words in modification. These represent relationships such as genitive, possession, part/whole, and association (16a-c).

(16) Noun-Noun modification constructions in Toqabaqita (Lichtenberk 2008b: 385, 379, 408)

a. Juxtaposition:
rua *araqi* *loo*
work mature.man upward
'God's work'

b. Personal suffix:
gwero-na *kuukua*
crest-3.PERS chicken
'a/the chicken's crest'

c. Associative suffix:
biqu-i *dudualinga*
house-ASSOC bee
'bee's nest/beehive'

It is not clear if these relationships are restricted to particular lexemes or subclasses of object words. This study does not aim to capture markedness differences among these kinds of modifying relationships.

4.5.6. Semantic shift

A classic example of semantic shift in my data can be found in Quechua. When property words are used in reference, they can appear without any structural coding. However, there is a recognizable shift in meaning.

(17) Objects and properties in reference in Quechua (Weber 1989:35-6)

a. *rumi-ta* *rikaa*
stone-ACC I.see
'I see a/the stone.'

b. *hatun-ta* *rikaa*
big-ACC I.see
'I see a/the big (one).'

In the second example, the resulting construction does not refer to the property of 'bigness', but rather someone or something that exhibits that property. Although this property word appears without any structural coding, it has taken on an object meaning in this construction (Croft 1991: 72-74, 2001: 70-5). This particular semantic shift for property words in modification is very common cross-linguistically. Any instance of semantic shift that changes the broad semantic class of a lexeme (e.g., from property to object) is considered too much semantic shift in this study.

Another example of unacceptable semantic shift is found for actions in reference. Some action concepts, such as COUGH and FIGHT, appear as Nouns in several languages. As Nouns, however, they refer to a product of the event rather than the event itself (Croft, pers. comm.). For FIGHT—like DANCE—the product is the structure of the action. For emission events, the often ephemeral product is what is emitted: sound, light, or substance. These products all represent non-prototypical objects rather than derived actions.

For some languages, specific derivational morphemes encode the relationship between the stem and the derived lexeme. This is at least generally true for the nominalization of Quechua's Verbs. The suffix –*q* is an "Agentive Substantivizer", by which the resulting Noun refers to the agent of the action denoted by the source Verb (Weber 1989: 53). Similarly, the suffix -*na* creates a Noun that refers to the tool used to do the action denoted by the source Verb, such as the derivation of 'broom' from the Verb meaning 'sweep' (50-1). Neither of these are relevant for this study, since they shift the referent from an action to an object. A third Quechua nominalizer is the suffix –*y*, an "Infinitive Substantivizer" (51-3). When attached to some Verbs, the resulting Noun appears to maintain its action meaning. However, other examples indicate a significant semantic shift on par with the other Quechua nominalizers discussed above. Still others are glossed in such a way to be somewhat ambiguous as to the degree of semantic shift.

(18) Infinitive Substantivizing suffix in Quechua (Weber 1989:51-3)

a. *miku-y*
eat-INF
'food'

b.

kanan	*papel-niki-kuna*	*tinku-chi-y-chaw*	*lloqshi-nki*	*alli*
now	paper-2.POSS-PL	equal-CAUS-INF-LOC	come.out-2	good

'Now in (the circumstance of) seeing if your paper measures up, you will come out good…'

c. *aqcha* *rutu-y* *ka-n*
hair cut-INF be-3
'There is hair-cutting going on.'

Because it is impossible to determine from the available data how this suffix interacts with each lexeme in this study, this information is left out of the analysis.

Kanuri features a similar semantic ambiguity for property words in reference:

> The word *cìmê* for example may be used as a modifier meaning 'red', or as a noun meaning 'the red one', or even to mean 'redness', though for the latter meaning there is also the alternative derived *nəm-cìmê* 'redness' formed through the affixation of the abstract nominal prefix… The word *kúrà* means 'big', 'important' as a modifier, and either 'boss', 'leader', 'head' or 'the big one', 'the important one', when used as an independent head… Any of the words that have traditionally been translated as and referred to as Adjectives in Kanuri can function as noun phrase heads, referring back to an anaphorically understood noun phrase antecedent. (Hutchison 1981: 36)

The semantically shifted meaning from property to object is generalizable to all members of the Kanuri Adjective class. However, it is unclear if all of them can render a property meaning in reference, as is described for 'red'. Since the latter meaning is the target of this study, some facts about how Kanuri's Adjectives are used in reference are simply missing from the analysis.

The examples presented above should make it clear that a primary challenge to the inclusion of derivational morphology in my analysis is the variability of semantic effect of some derivational morphemes. Combined with different lexemes, the same morpheme can often derive results with critically different levels of semantic shift. This variability makes it very difficult for the grammar writer to make useful generalizations without compromising some of the details that are so crucial for this kind of study.

This topic also has implications for the overall prototype structures of parts of speech. I will demonstrate in later sections how the full range of markedness data supports many of my hypotheses about which semantic criteria are most influential in shaping the prototype structures. Some of these are semantic features that exert their influence at the hearts of the prototypes. For example, the animacy hierarchy most commonly creates distinctions in inflectional potential within the prototypical combination of objects in reference. These derivational constructions, however, are primarily affecting structural coding in non-prototypical combinations of semantic class and propositional act function. Interestingly, the same semantic criteria that I've predicted to be relevant at the prototypes and overall may be different from the criteria most relevant to minor distinctions in the "no man's lands" where these derivational constructions are at work. This is because semantic shift becomes the dominating factor in these non-prototypical constructions.

When the semantic class does not "match" the propositional act function, there is a tug-of-war between the semantic contribution of the lexeme and the semantic contribution of the construction. This is why, for example, property words in reference so often semantically shift to refer to a person or thing that exhibits that property. The newly formed object word fits comfortably in reference, as objects generally do. The natural fit and frequency of this semantically shifted derivational construction may coerce a language into deriving a different construction for the less-commonly used reference of the property itself, and that alternative construction may be marked relative to the derivation involving semantic shift. For those languages in which the derivational semantics are inconsistent and lexically specific, markedness patterns could boil down to how predictable a semantic shift is for that lexical concept. As discussed above, some action concepts refer to an event of emission with an identifiable product. The product may be an obvious target for semantic shift when that lexeme is used in reference, while other action concepts without such an obvious target may resist a similar semantic shift. This would result in differences in markedness for these actions in reference. The semantic characteristics that determine whether an action has an obvious target for semantic shift may be different from the aspectual characteristics that were hypothesized to be relevant for this study. This could be investigated with more detailed language data, particularly for those languages in which only some action words undergo semantic shift when used in reference.

4.5.7. Apparent arbitrariness

The examples from the data and the discussion of challenges to the methodology have illustrated that markedness differences arise out of the confluence of several factors. Historical developments interact with frequency effects and processes of morphosyntactic reanalysis and phonological fusion. A comprehensive understanding of these factors would require consideration of the unique circumstances in each language and in the evolution of each construction. For example, Creek Verbs (this class includes Adjectives) index the number of their subject (intransitive) or object (transitive) in three distinct ways (Martin 2011:197). A small set of common Verbs (e.g., ‘sit’, ‘come’, ‘die’, ‘break’) achieve this through suppletion. Alternatively, reduplication accomplishes this inflection for many stative

Verb roots (such as Adjectives—e.g., 'white', 'soft'). Still another subset of Verbs (e.g., 'red') utilize a suffix. These different strategies obviously reflect the influences of frequency and stativity, but the full story is much more complex. How does 'red' end up in a different class than 'white'? At least on the surface, frequency and stativity do not appear to answer this question.

In Kanuri, seemingly minor differences in the phonological form of a Verb turn out to have a major impact on markedness (Hutchison 1981:157). Verbal Nouns (basically infinitives) can be derived from Class 1 Kanuri Verbs in two different ways. For CV roots featuring any vowel except /i/, a tone change accomplishes this derivation (e.g., *tá-* > *ta*). In contrast, the suffix *–o* is added to derive the Verbal Noun for almost all CV roots with /i/, CVC roots, and polysyllabic roots (e.g., *dí-* > *di-o*, *lad-* > *lad-o*). This suffix is additional structural coding in my analysis, and the resulting Verbal Noun form is marked relative to those Verbal Nouns derived from only a change in tone. The vowel of a Kanuri CV root is not truly random, since it comes from a particular set of historical processes at work since its inception. Yet it seems to be a fairly trivial characteristic of the root from which a significant markedness distinction has arisen.

These examples are illustrative of the myriad of factors motivating the prototype structures of parts of speech and the complex relationships among these factors. That a recognizably constrained pattern emerges from these complexities is a testament to the strength and universality of the functional pressures motivating parts of speech.

5. Results and discussion

The multidimensional scaling analysis was successful in representing most of the variation in the data in only two or three dimensions.

Table 10. Fitness statistics in one, two, and three dimensions for the MDS analysis of parts of speech

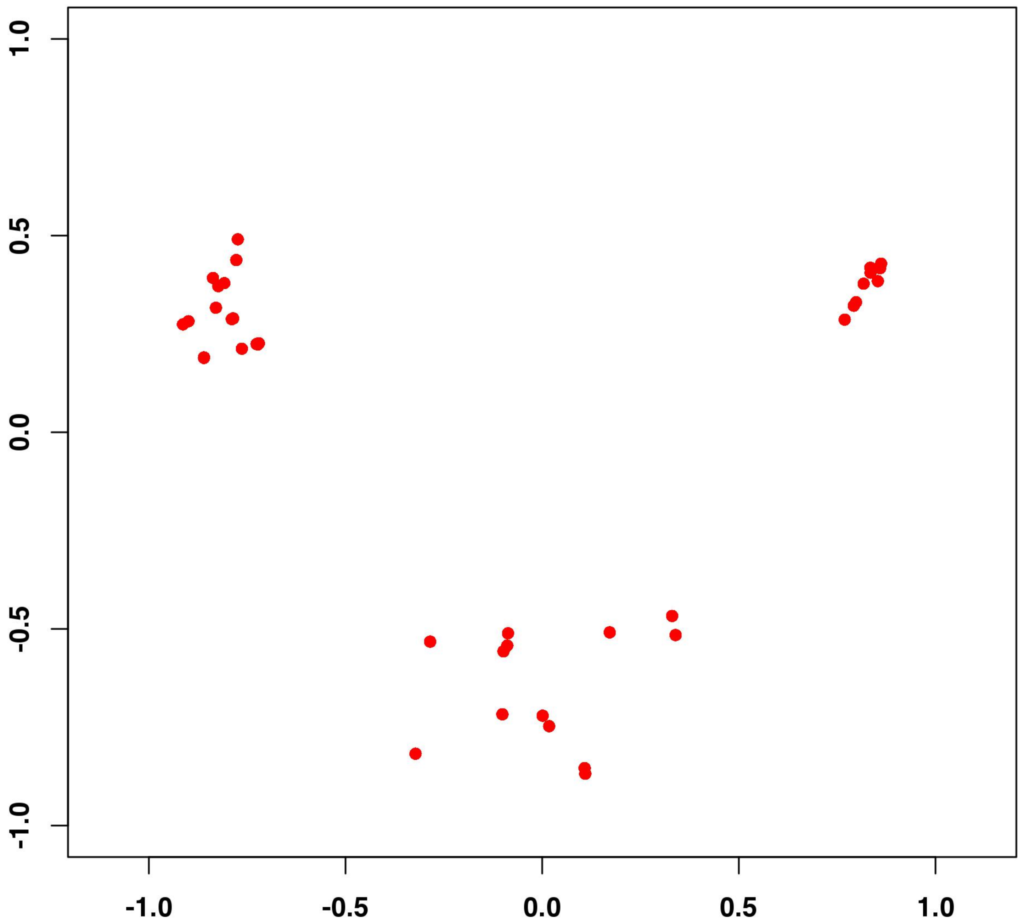
Dimensions	*Classification*	*APRE*
1	0.94030	0.79240
2	0.98679	0.95406
3	0.99212	0.97261

As discussed above, the best model yields a high degree of fit, and for which each additional dimension offers only small improvements to the fit. For the parts of speech data, a two-dimensional model clearly fits this description. The classification and especially the aggregate proportional reduction of error show a significant improvement from one dimension to two. However, from two to three dimensions, these numbers only show small improvement. The two-dimensional model demonstrates an excellent fit while remaining very informative—it characterizes a large data set with very low dimensionality.

A common result of an MDS analysis in two dimensions is a horseshoe shape (Borg and Groenen 1997, cited in Croft & Poole 2008: 17), and this shape was found for the parts of speech data. The data are arranged linearly, and that line is bent into a horseshoe shape allowing straight cutting lines to isolate smaller strings of data that fall along the line. The linear nature of this data corresponds to the semantic continuum from objects to properties to actions. If some grammatical functions apply to a subset of the data points (lexemes) in the middle of the linear arrangement—such as an Adjective category in a language that includes properties but excludes objects and actions—the line must be curved to allow a straight cutting line to separate those points from the rest. Figure 3 shows that the horseshoe shape points upward; i.e., the open end of the horseshoe is at the top of the figure. For more evenly distributed data, points will appear all along the horseshoe shape with the ends drawing somewhat near to each other. Because the parts of speech data are dominated by three clusters corresponding to Nouns, Adjectives, and Verbs, the ends of the horseshoe shape do not come together in this way. An open end still faces upward, but the points form more of an open, "U" shape.

There is another important characteristic of this horseshoe shape that will be relevant in the discussion below. Although the data points are generally aligned in the shape of the horseshoe, some points end up closer to the inside of the horseshoe and some end up along the outer edge of the horseshoe. This is not an accident, but rather an emergent characteristic of the spatial model. The points along the outside of the horseshoe can more easily be isolated from the rest by a cutting line, while points along the inside of the horseshoe cannot. Therefore, if grammatical distinctions separate one particular conceptual target from the rest, that point is likely to show up along the outside of the horseshoe. Conceptual targets that are never or less-often individuated in the data will likely end up along the inside of the horseshoe. In this way, the position of a point along the inside or outside of the horseshoe is theoretically significant, not just its position along the curved line from one end of the horseshoe to the other.

Figure 3. MDS analysis of parts of speech with points only



5.1. Distribution of the prototypes

The prototype structures of parts of speech corresponding to Nouns, Adjectives, and Verbs are immediately clear in the spatial model. Before investigating each of these in more detail, I first present some observations that may seem basic but nevertheless have theoretical importance.

Figure 4. MDS analysis of parts of speech with cutting lines only

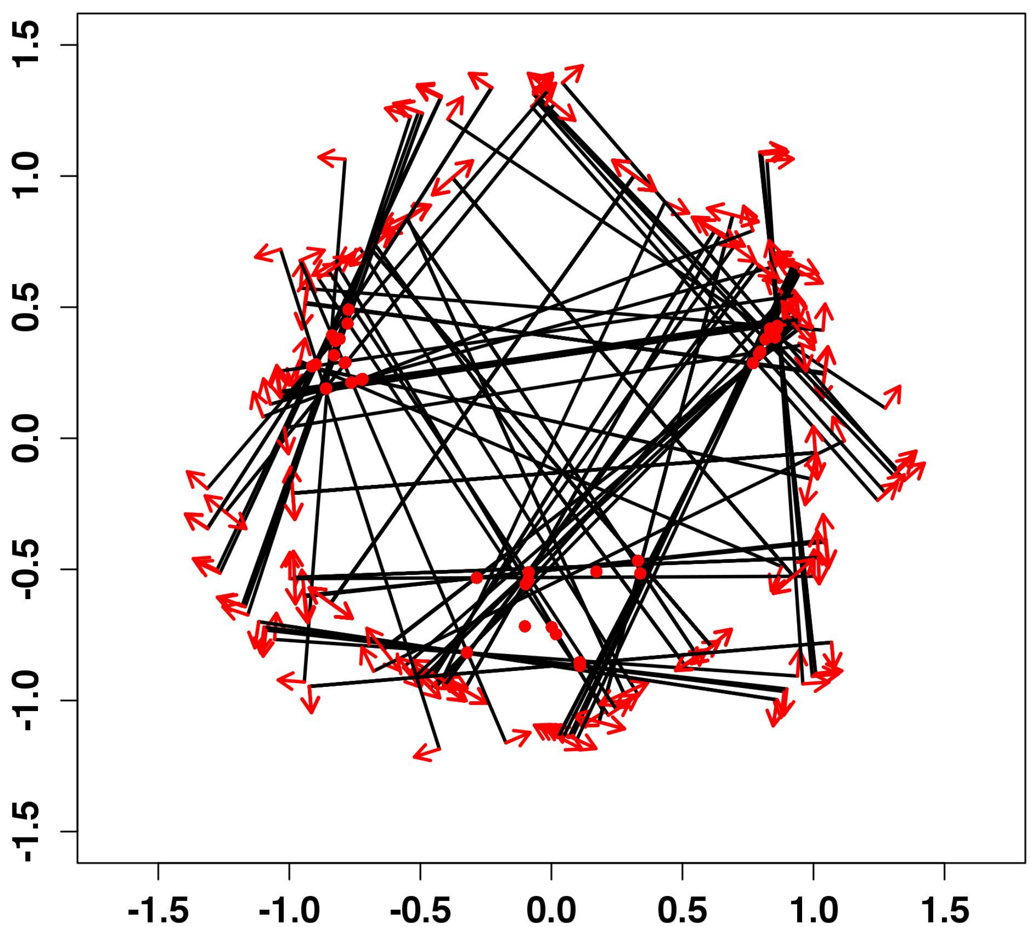
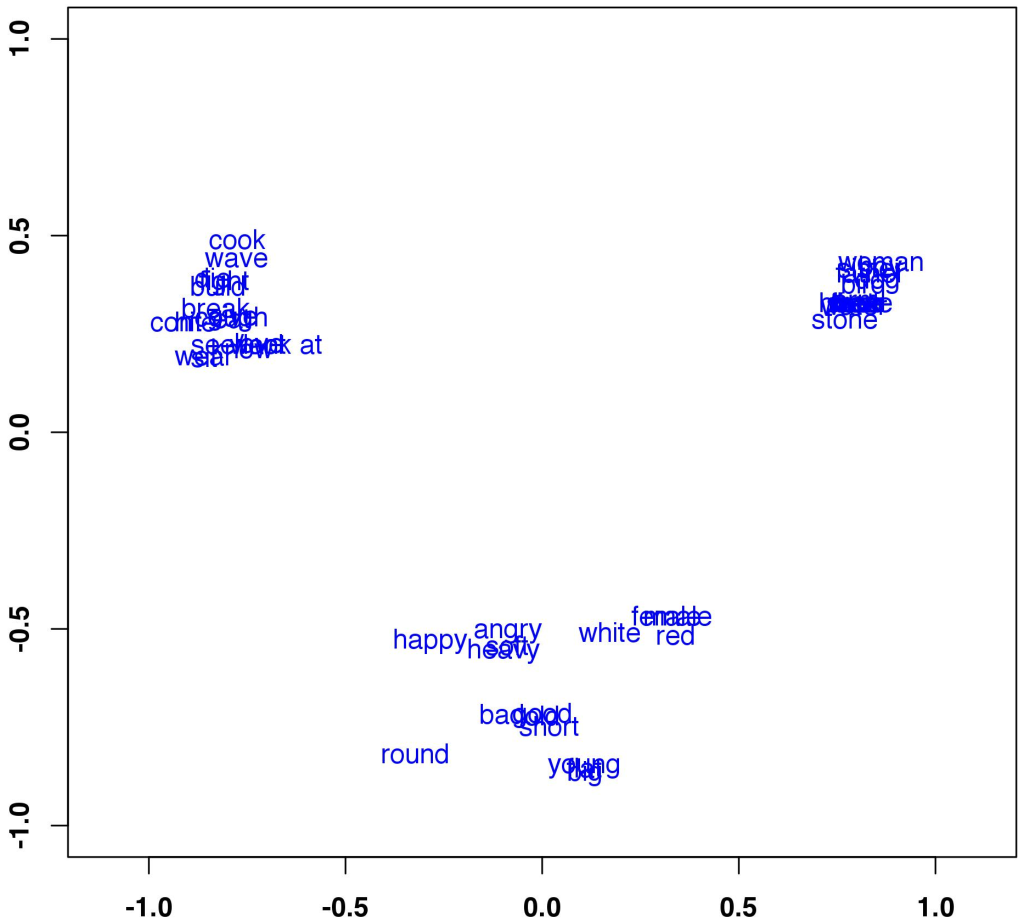


Figure 5. MDS analysis of parts of speech with concept labels only

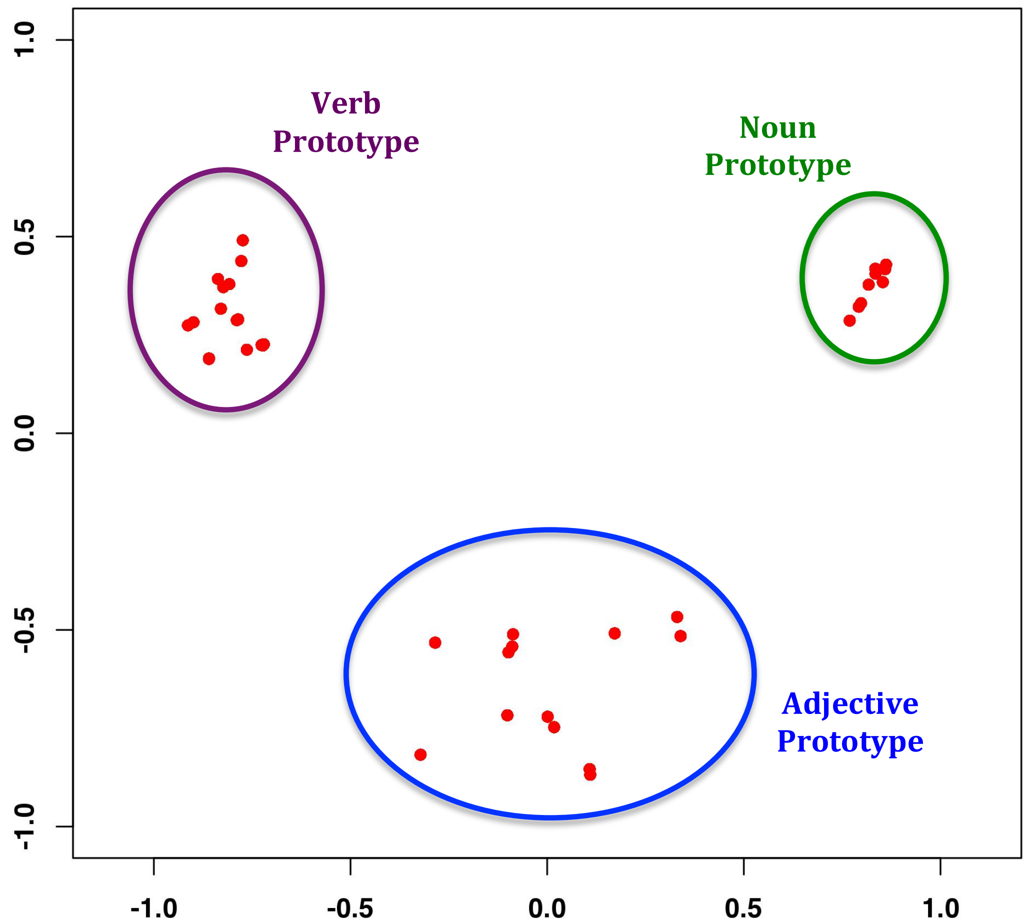


There are some points that are directly overlapping, meaning there were no distinctions among them in the data. Yet many of the data points are differentiated in the analysis. Why is this important? First of all, there were many conceptual targets even within each broad semantic class. Each of these broad classes—objects, properties, and actions—included at least 16 targets. Some of the conceptual targets fall into the same subclass, such as the age concepts YOUNG and OLD. Yet almost all of these conceptual targets were differentiated from each other by a markedness distinction in at least one of the eleven languages. Languages exhibit considerable variation regarding where they make meaningful, grammatical distinctions along the objects-properties-actions semantic continuum.

Despite the differentiation of nearly every data point, the broad classification of these concepts into objects, properties, and actions is supported in the spatial model. There are greater dissimilarities across these broad classes than within each one. Every language in the study makes significant markedness distinctions between most objects and most properties, and between most properties and most actions. Languages make some different decisions in their classification of conceptual targets that are close to the boundaries between these broad semantic classes, but they all make at least one major distinction near those boundaries. In fact, this is the primary reason for analyzing each of the three clusters as separate prototypes.

I now turn to the prototypes themselves. First, strong noun and verb prototype structures can be observed in the upper right and upper left regions of the analysis, respectively. As discussed in section (2.6.2.), the spatial model is based on *dissimilarity* data, so prototype clusters are an indirect result of greater and lesser levels of dissimilarity among points. The discussion here of “stronger” prototypes refers to data point clusters which exhibit less dissimilarity, resulting in points that are closer together in the analysis. In addition to the noun and verb prototypes, a weaker adjective prototype is observable in the lower, middle region of the analysis. These data points represent property concepts, and while they exhibit more dissimilarity among them than is found in the other prototypes, they maintain some amount of cohesion as well. These clusters are illustrated in Figure 6.

Figure 6. MDS analysis of parts of speech with points and prototypes



In the next sections, I turn to a discussion of the internal structure of each prototype. The arrangement of points in the spatial model will be evaluated to determine whether they substantiate my predictions regarding the relevant semantic characteristics for prototypicality.

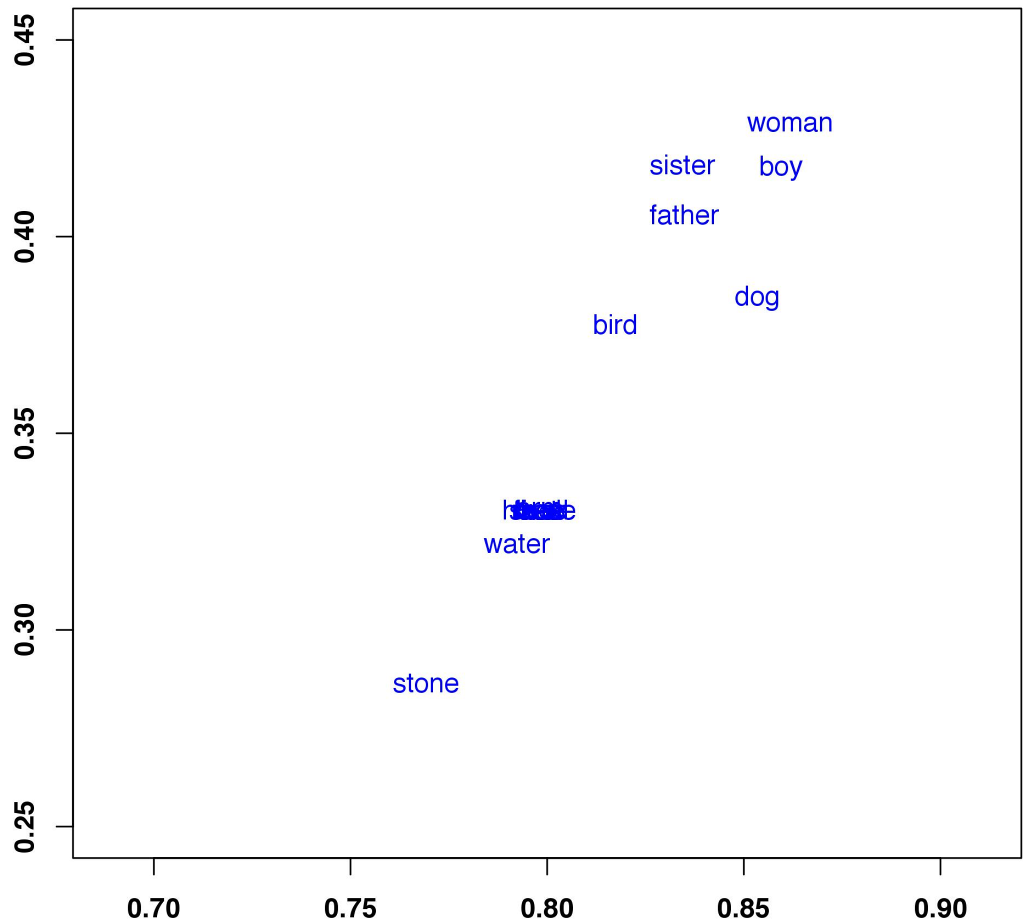
5.2. Noun prototype

The noun prototype is probably the most straightforward in both the spatial arrangement of its data points and the semantic motivation for this arrangement. Figure 7 offers a closer look at the data points (represented by their labels) that constitute the noun prototype. All these points represent object concepts, and they nearly form a straight line in a steep, positive-sloping diagonal. Recall that a curved line is necessary for cutting lines to separate medial segments from the rest. The object concepts are generally aligned on a straight line because there is one primary functional motivation that dominates their arrangement: animacy.

Animacy was responsible for several markedness distinctions across more than one language in this study. These included human/non-human, higher-animate/lower-animate, and animate/inanimate distinctions. As a result, the uppermost points in this cluster of the spatial model are the object concepts that are most animate. All four human concepts are found at this extremity, and these are followed by non-human animates—first DOG and then BIRD. WOMAN was also differentiated from BOY even though they are both human animate concepts. For example, the English ablaut plural form *women* is less marked than the suffixed plural form *boy-s*.

At the lower end of the objects cluster, STONE and WATER are stretched out in the direction of the adjective prototype. Lexemes for STONE patterned with property words in several languages, which explains its spatial orientation. WATER is unable to inflect for number in Kisi due to its status as a mass noun, distinguishing it from other object concepts in this analysis. This suggests that unbounded objects are most distant from the noun prototype among object concepts.

Figure 7. Noun prototype



Between these two concepts on the lower end and the animates on the upper end, a significant number of concepts landed in the same spot in the spatial model. This group includes TREE, SEED, HAT, BED, ARM, FACE, RIVER, and HOUSE. No markedness distinctions were found to disambiguate these terms in my analysis. The explanation may be found in the semantic features that were used to subclassify these object concepts in the first place. These concepts represent several subclasses: plants (and plant products), artifacts, body parts, and places. They are not as readily ranked in terms of animacy. Body parts (includes parts of humans, animals, and plants) are indirectly based on animacy by the objects' relationships with humans or other animates. Places are subclassified separately from artifacts based on magnitude (among other things), especially relative to how humans experience them. It is possible that distinctions in markedness would be found to distinguish these subclasses and their concepts if more languages were added to the study. Even so, this study suggests that plants and body parts are treated like inanimates with regard to parts of speech. Distinctions among these concepts might be reserved to lower level grammatical structures. For example, alienable and inalienable possession relationships are encoded differently in many languages, which could disambiguate body parts from artifacts. This grammatical distinction is less obviously related to the occurrence of these lexemes in a particular propositional act function; rather, it has implications for how that object concept takes part in a possessor-possessee relationship.

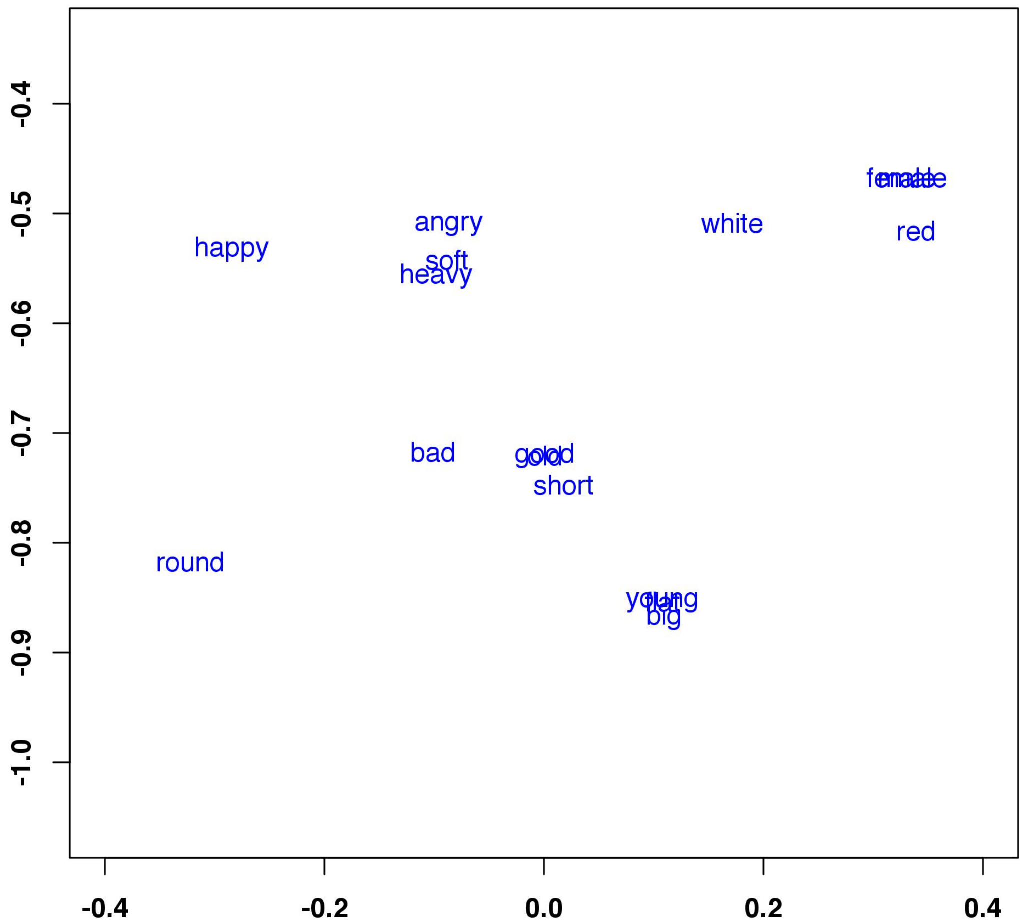
In summary, the influence of animacy is robust in the noun prototype in this analysis. Unlike the other prototype clusters discussed below, the object concepts are nearly arranged in a straight line reflecting this implicational hierarchy. Furthermore, some known conceptual distinctions such as alienable/inalienable and plant/inanimate may only be relevant for grammatical structures unrelated or indirectly related to parts of speech.

5.3. Adjective prototype

The most obvious characteristic of the adjective prototype is that it is less cohesive than its noun and verb counterparts. These property concepts form a relatively loose cluster that covers a larger area of the spatial model. Furthermore, this cluster does not formed the curved line that we might expect at the bottom of a horseshoe arrangement. Instead, the points are spread significantly to the inside and the outside of the horseshoe. Figure 8 gives a closer look at this section of the spatial model. A few overlapping concepts need to be identified, since they are difficult to read even in the enlarged image. The overlapping concepts in the upper right corner of this section are MALE and FEMALE. In the center of the section, just above SHORT, are the concepts GOOD and OLD. Finally, YOUNG, FLAT, and BIG are the at least partially overlapping concepts in the lower right section.

The analysis of property concepts in the literature as a continuum from object-like to action-like is partially supported by the analysis. Specifically, those property concepts that have been placed on the extremes of this continuum show up this way in the analysis. The gender concepts MALE and FEMALE were expected to be the most object-like property concepts based on this literature, and they appear in the upper right corner of the property concept cluster, closest to the noun prototype. As predicted, the color concepts WHITE and RED appear in the upper right corner as well, just below and to the left of the gender concepts. On the action-like side of the continuum, human propensities are found in the upper left corner of this cluster, closest to the verb prototype. Just a little closer to the center of the cluster, physical property concepts SOFT and HEAVY are right behind. These observations in the spatial model reinforce the claims in the literature.

Figure 8. Adjective prototype



The eight property concepts that are in the middle of the continuum—neither very object-like or action-like—are not arranged so neatly in the spatial model. Some of these concepts are clustered very close to each other, and there does not seem to be an obvious structure from "nouny" to "verby" within them. This may result from the limited data set used in this study. Data from additional languages would surely tease out the details at the heart of this prototype structure. A couple of observations will have to suffice for this paper. First, ROUND appears in the lower left corner of the cluster, with a significant amount of separation from the rest of the property concepts. It is not clear what motivates this position, or whether the inclusion of additional language data would perpetuate its outlying position in the analysis. Second, BIG appears in a low, relatively center position in the properties cluster, and therefore on the outside of the horseshoe shape. This position reflects the status of BIG (and other common concepts representing dimension, such as LITTLE) as an extremely prototypical property concept which enjoys unique inflectional capabilities in modification in some languages.

What are the semantic motivations behind the prototype structure that I have just described for property concepts? Although the literature identifies "nouny" and "verby" property concepts, the semantic motivations for these characteristics are not always explicated. Based on the markedness differences observed in this study, I will discuss the semantic motivations that seem to be influencing the structure of the adjective prototype in the spatial model.

The concepts MALE and FEMALE appear in the upper right corner of the adjective prototype, closest to the noun prototype. In several languages of this study, these concepts are expressed with lexemes classified as Nouns. As a result, these lexemes are unable to participate in Adjectival inflections, such as comparatives and superlatives. This may reflect a construal of these concepts as lacking gradability, and further evidence comes from languages which encode these concepts as

Adjectives. In English, *male* and *female* cannot take comparative and superlative suffixes like many other property concepts; instead, they inflect periphrastically with *more*/*most*. Similarly in K'ichee', these concepts cannot inflect for comparative like other property concepts, despite being classified as Adjectives based on other formal criteria.

Table 11. Semantic properties of prototypical parts of speech (Croft 2001: 87)

	Relationality	*Stativity*	*Transitoriness*	*Gradability*
Objects	nonrelational	state	permanent	nongradable
Properties	relational	state	permanent	gradable
Actions	relational	process	transitory	nongradable

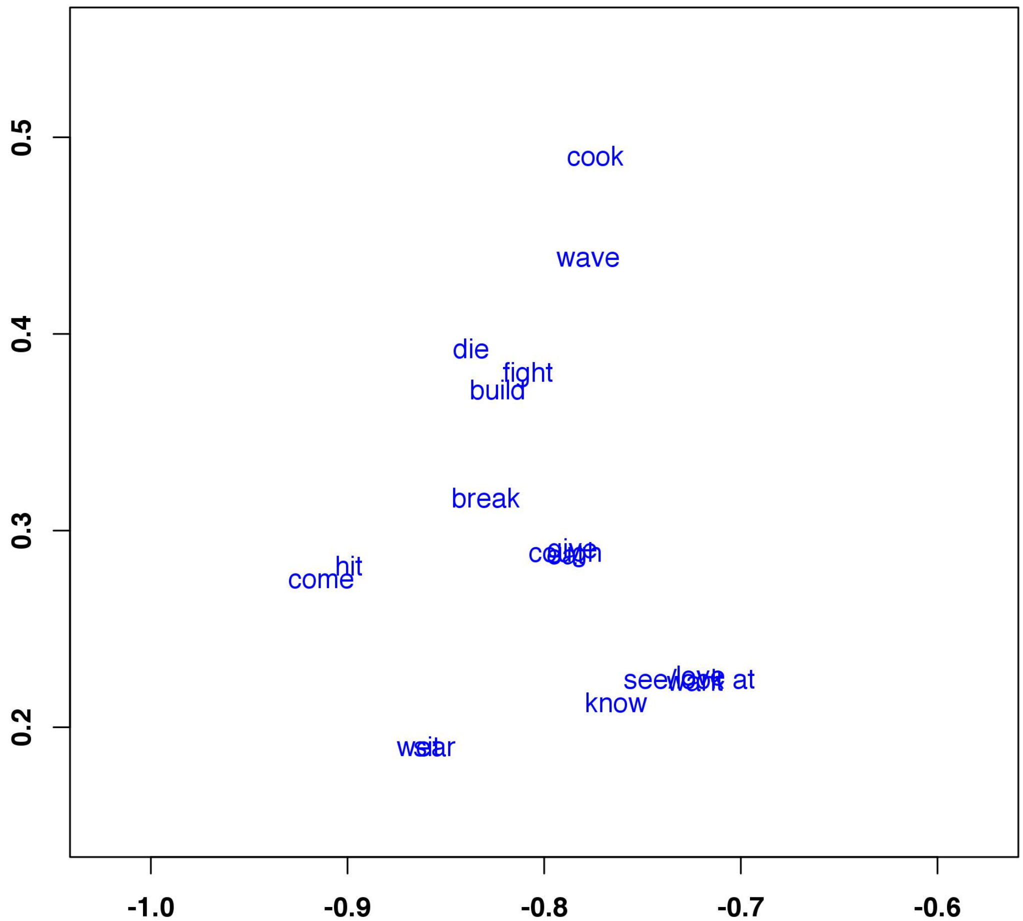
On the "verby" side of the property concept continuum, the semantic primitives that were claimed to distinguish properties from actions are gradability, stativity, and transitoriness. Gradability does not seem to play much of a role here. In fact, some of the concepts that landed in the verb prototype cluster can be construed as gradable, such as WANT and LOVE. In contrast, the role of stativity and transitoriness can be seen in the most "verby" property concepts: those of human propensity. First of all, HAPPY and ANGRY are known through human experience to be transitory. At least in English, they can refer to an overall personality characteristic, but more prototypically they evoke a temporary attitude or demeanor. This transitoriness of human propensities is tied up with their construal as a process, since humans have to transition into and back out of a particular temperament. In fact, the various possible construals of an experience of human anger present a challenge to this study. A lexeme recruited to represent the experience might target the resulting state or the inchoative process, for example. It is precisely these semantic characteristics inherent to the experience of human propensity that lead to its occasionally "verby" morphosyntactic treatment.

5.4. Verb prototype

The action concepts form a relatively tight cluster in the spatial model, although not quite as tight as the noun prototype. Also, as is found in the adjective prototype, the actions concepts not only follow the curved line of the horseshoe but exhibit variation to the inside and outside as well. This is indicative of a more complex set of semantic motivations behind the morphosyntactic realization of these concepts. In the lower left, the overlapping concepts are SIT and WEAR. In the lower right, SEE/LOOK AT, LOVE, and WANT are bunched together. Finally, GIVE, COUGH, and EAT are the overlapping concepts near the center of Figure 9.

As discussed in 2.2.3, aspectual distinctions were expected to play a role in the prototypicality of action concepts. This hypothesis was confirmed with regard to one particular distinction: stative concepts versus dynamic concepts. Two-participant states (KNOW, WANT, LOVE) and inactive actions (SEE/LOOK AT, SIT, WEAR) are stative concepts, and these are clearly separated from the rest of the action concepts in the spatial model. As expected, they are found in the lower and lower right areas of the verb cluster, closer to the adjective prototype than the rest of the action concepts. As discussed in some examples above, the stative/dynamic distinction was found to be significant for markedness patterns in some of the languages in this study, and the results are apparent in the model.

Figure 9. Verb prototype



Within this subgroup of stative concepts, however, the distinction between two-participant states and inactive actions is not clear in the spatial model. Inactive actions SIT and WEAR are found together in the lowest position in the verb cluster, but SEE/LOOK AT—also an inactive action—is bunched up with the two-participant states . The separation of SIT and WEAR from the other stative concepts may reflect the unintentional inclusion of more active meanings in some of the lexemes chosen to represent these concepts. For some languages, the same lexeme can mean 'sit' and 'sit down' or 'wear' and 'put on'. These alternative meanings—often difficult to disentangle from the stative meanings—might be motivating these lexemes to pattern morphosyntactically in some ways with dynamic action concepts.

The remaining action concepts are spread from the center and left center sections of the cluster toward the upper right. This generally follows the curved line of the expected horseshoe shape ending at COOK. The semantic motivation for the arrangement of these dynamic action concepts in the model is less clear. Durative processes, cyclic achievements, and directed achievements are all intermingled together in this part of the verb prototype.

There is, however, a tendency suggestive of a durative/punctual distinction.[7] Durative concepts tend to be found at the upper and upper right portions of the verb prototype. This grouping might include COOK, WAVE, DIE, FIGHT, and BUILD. WAVE was chosen for this study to represent a cyclic achievement, but iteratively construed, it could take on a durative sense. DIE would be a clear exception to the durative theme. Punctual concepts, on the other hand, are found clustered in the center and left center of the prototype. These include COME, HIT, BREAK, GIVE, COUGH, and EAT. COME was selected to represent a durative process, but an alternative construal for this concept—'arrive'—could explain its place among punctual concepts. However, it is difficult to conceive of a strictly punctual construal of EAT, so this concept must be considered an exception in this group.

---

[7] I am thankful to Bill Croft for bringing this to my attention in our discussion of the results of the MDS analysis.

If we accept the durative/punctual distinction suggested in the model, why are the durative concepts found at the end of the horseshoe (upper right portion of the action concepts cluster) where verb prototypicality is highest? A hypothesis of this study was that punctual actions—represented by cyclic achievements and directed achievements—would be more prototypical among actions than durative ones. The answer may lie in an unexpected relationship between achievements and states. From a perspective of boundedness or completion, these two aspectual classes appear to be very different. However, there is evidence that they are united as potential construals of a directed aspectual contour (Croft 2012: 165-71). This contour corresponds to one of Bybee et al.'s (1994) families of grammatical tense-aspect categories which includes perfect and perfective senses. Synchronically and diachronically, verbs can alternate between an achievement construal and (resulting) state construal. (Alternatively, durative concepts are construals of an undirected aspectual contour, corresponding to Bybee et al.'s other family of grammatical tense-aspect categories which includes imperfective, progressive, present, and habitual senses.) This surprising relationship may be motivating the contiguity of punctual and stative concepts in the parts of speech model.

The spatial model has provided various insights into the nature of the verb prototype, but it does not have to explain every aspect of the data. Several of the languages in this study feature important markedness distinctions among action concepts—particularly with regard to how they are predicated—that do not seem to result from aspectual distinctions. For some languages these distinctions are the result of verb classes, for other languages it is due to the existence of semantically meaningful auxiliary verbs. It is also possible that individual lexical frequency manifests itself in morphosyntactic markedness enough to disrupt the influence of aspectual prototypicality. It is interesting that this does not also happen within object concepts. For example, not every lexeme representing a human concept is frequent, yet in certain languages the morphosyntactic “advantages” such as inflectional potential are applied to all lexemes representing human concepts. What differentiates the action concepts from the object concepts in this regard? It may be that the conceptual unity of subclasses in the animacy hierarchy is greater than that of aspectual subclasses. Alternatively, the grammatical processes by which markedness patterns evolve in objects may be different that those for actions, so it may be only indirectly related to the semantic characteristics that motivate these subclasses.

5.5. Interpreting the dimensions

Parts of speech represent one of the most basic and pervasive levels of grammatical organization, and it would be extremely surprising if only two or three functional or semantic factors were shown to explain nearly all of the data. Rather, these dimensions represent different semantic motivations in different areas of the spatial model. For example, within the noun prototype cluster, one could argue that an animacy dimension runs along a steep, positive-sloping line; highly animate concepts are higher and further to the right, while inanimate concepts are lower and further to the left. However, an animacy interpretation of this dimension is inappropriate for other areas of the model where other semantic factors appear to be driving the distribution of concepts.

The multidimensional scaling analysis confirms the hypothesis of this study that at least several semantic factors are relevant for the morphosyntactic realization of objects, properties, and actions in three propositional act functions. Table 12 shows the primary semantic factors based on where their effect is found in the semantic continuum from objects to properties to actions. Most of these semantic primitives are familiar from Croft (2001: 87, see also 1991), and they were introduced in Table 1. However, the effect of animacy in this analysis is strong enough to include it with the other semantic factors.

Table 12. Semantic properties motivating the prototype structures of parts of speech

<table>
<tr><td colspan="3">ACTIONS</td><td colspan="2">PROPERTIES</td><td colspan="3">OBJECTS</td></tr>
<tr><td>dynamic</td><td colspan="7">stative</td></tr>
<tr><td colspan="3">transitory</td><td colspan="5">permanent</td></tr>
<tr><td colspan="2">nongradable</td><td colspan="2">gradable</td><td colspan="4">nongradable</td></tr>
<tr><td colspan="5">relational</td><td colspan="3">nonrelational</td></tr>
<tr><td colspan="6">inanimate</td><td>animate</td><td>human</td></tr>
</table>

While Table 12 implies that all of these semantic primitives are active across the entire continuum from objects to properties to actions, this may not be the case. Stativity and transitoriness are components of the larger concept of time-stability, and they appear to be relevant across the continuum. Similarly, relationality motivates a distinction between objects and properties, so its effects are not confined to a single prototype. However, animacy may be relevant only among nonrelational concepts, effectively limiting its domain to the noun prototype. It may be appropriate to say that animacy contributes to the internal structure of the prototype, but not to the (non-discrete) boundary between prototypes. It is less clear, but gradability may be similarly limited in its scope. Its effects are primarily found in the middle of the continuum—that is, among property concepts—and less so on either end.

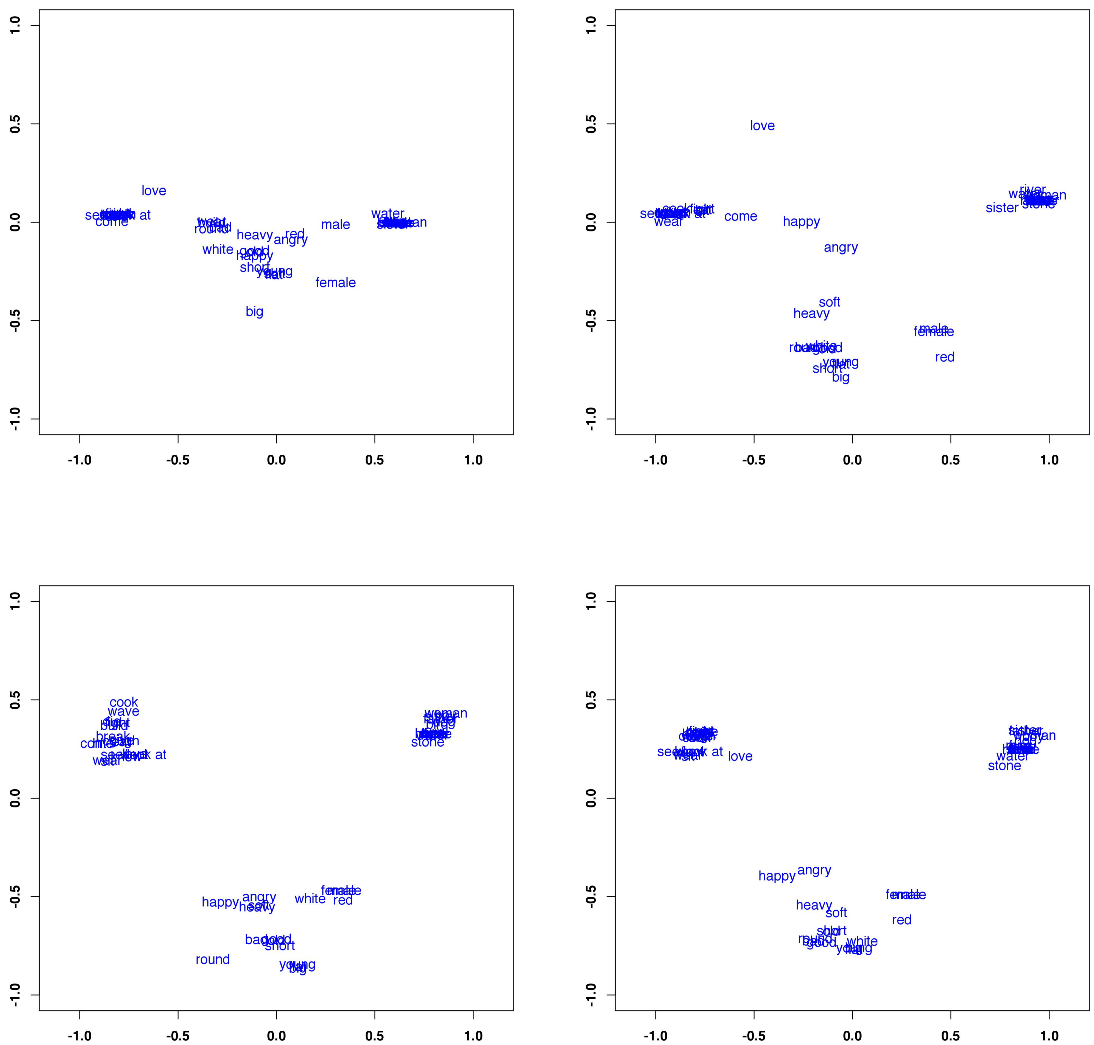
The analysis presented here shows that humans can indeed reduce the complexity of the world into just a few dimensions, but that those dimensions may actually represent various functional or semantic principles simultaneously.

5.6. Future research

The most obvious extension of this study would be its application to more languages. The eleven languages included in this study are just enough to illustrate the prototype structures. This can be seen in Figure 10, which shows spatial analyses using data from three, six, nine, and eleven languages. It takes close to a dozen to tease apart most of the concepts in the spatial model. Doubling or tripling the sample size would go a long way in further elucidating dissimilarities and their semantic motivations.

Additionally, future research can improve on the list of concepts included in such a study. Of course, with more fine-grained conceptual distinctions, a more informative model is possible. However, economical considerations limit the number of concepts that can be included, so future work might instead replace uninformative conceptual distinctions with ones that are predicted to be relevant based on the results here.

Figure 10. MDS analysis of parts of speech for three, six, nine, and eleven languages (clockwise from top left)



5.7. Experience, conceptual structure, and linguistic form

This study is an investigation into how the same conceptual idea may be encoded grammatically in different ways from one language to the next. This is the source of variation between the parts of speech prototypes. There is, however, another source of variation in parts of speech among languages. The same real world target—whether an object, property, or action—can be *construed* differently across languages. This, of course, would result in different morphosyntactic behaviors for the lexemes representing these different construals. For example, two languages may construe the experience of anger differently, with one focusing on the temporary state of being angry and the other focusing on the process of change from a non-angry state to an angry one. A study that starts at a conceptual baseline misses this kind of variation. This thesis aimed to capture variation starting from a conceptual baseline, but as we have seen, alternative construals fossilized in lexical polysemy have most likely introduced some of this variation in construal.

There are many studies that have aimed to capture the variation in construal, perhaps most notably the pioneering work of the Pear Stories (Chafe 1980). The use of video allowed participants in the study to go directly from experience to linguistic form, without a predetermined syntactic or conceptual structure. For some targets in the film, the linguistic results are surprisingly homogeneous; for others, a great deal of variation resulted. It would be interesting to conduct a study on parts of speech from an experiential baseline. This would require video or picture elicitation, and a great deal of time and resources to collect the data from many diverse languages. The result would likely be more variation than is found in this study, but how much more? Would the parts of speech prototype structures emerge in such a study, or would they be hidden by the confounding effects of variation in construal?

6. Conclusions

This thesis has sought to illustrate the prototype structures of parts of speech. After coding markedness asymmetries in eleven languages, I was able to apply a multidimensional scaling analysis to the data to create a two-dimensional spatial representation that captures almost all of the variation. The model shows clearly three prototype structures representing nouns, adjectives, and verbs that result indirectly from conceptual dissimilarities. The noun and verb prototype clusters are stronger, while the adjective prototype cluster is weaker.

In particular, it is the interaction of conceptual classes and subclasses with the three basic propositional act functions that yield markedness differences. Languages exhibit variation regarding which concepts belong to which language-specific classes and regarding how these concepts are encoded morphosyntactically. Yet through this variation, the prototypicality of objects in reference, properties in modification, and actions in predication motivate the structures that are seen clearly in the spatial model presented here. It is significant enough to emphasize again that every language in this study was found to exhibit some morphosyntactic distinctions between objects, properties, and actions. This characteristic of the data is critical in motivating a spatial model with three distinct clusters, rather than continuous and inseparable data points along the semantic continuum from objects to properties to actions.

One of the most important contributions of this study is to our understanding of the semantic factors that motivate the parts of speech prototypes. The role of several semantic factors identified in the literature were supported in this analysis, such as relationality, gradability, transitoriness, and stativity. The evidence for their role is seen both qualitatively in markedness distinctions of particular languages that reflect their influence and quantitatively in the arrangement of concepts in the spatial model. Animacy has also been discussed in the literature as shaping formal distinctions among lexemes representing object concepts, but was not included as a primary semantic factor in parts of speech prototype structures. The analysis here has shown that it plays a key role, but also that this role may be confined to the internal shape of the noun prototype. This opens the door for future research to explore the domain of each semantic factor—within a particular prototype, shaping the boundary between two or more prototypes, or both.

These results support and illustrate the prototype theory of parts of speech. Even more, they are consistent with contemporary typological research and specifically multidimensional scaling analyses that support neither an "extreme universalist" nor an "extreme relativist" perspective on language (Croft & Poole 2008: 31-2). The variation in parts of speech strategies cannot be explained in terms of universal categories, yet the patterns that emerge are indicative of universal functional pressures that constrain the variation probabilistically. The spatial model of parts of speech data presented here illustrates the universal conceptual structure underlying the most basic organization of grammar.

LIST OF APPENDICES

Appendix A. Complete concept subclassification and selected exemplars

Class (/Type)		***Exemplars***
Humans		WOMAN
		BOY
Kinship		FATHER
		SISTER
Animals		DOG
		BIRD
Plants/Plant Products		TREE
		SEED
Artifacts		HAT
		BED
Body Parts		ARM
		FACE
Places		RIVER
		HOUSE
Material/Substance		WATER
		STONE
Gender		MALE
		FEMALE
Color		WHITE
		RED
Form		ROUND
		FLAT
Age		OLD
		YOUNG
Value		GOOD
		BAD
Dimension		BIG
		SHORT
Physical Properties		HEAVY
		SOFT
Human Propensity		HAPPY
		ANGRY
2-Participant States	Cognition	KNOW
	Desire	WANT
	Emotion	LOVE
Inactive Actions	Perception	SEE/LOOK AT
	Posture	SIT
	Spatial/Possession	WEAR
Durative Processes	Social Interaction	FIGHT
	Consumption	EAT
	Motion	COME
	Change of State	COOK
	Creation Verbs	BUILD
Cyclic Achievements (Semelfactives)	Emission	COUGH
	Contact	HIT
	Bodily Movement	WAVE
Directed Achievements	Existence	DIE
	Change of Possession	GIVE
	Change of State	BREAK

Appendix B. Parts of speech data coded for MDS analysis

	KH-R-SC-0	KH-R-SC-VarStr	KH-R-BP-PGN	KH-R-BP-0	KH-M-SC-0	KH-M-SC-VarStr	KH-M-BP-SPRL1	KH-M-BP-SPRL2	KH-M-BP-0	KH-P-SC-0
WOMAN	1	6	1	6	6	1	6	6	1	6
BOY	1	6	1	6	6	1	6	6	1	6
FATHER	1	6	1	6	6	1	6	6	1	6
SISTER	1	1	1	1	1	1	6	1	1	1
DOG	1	6	1	6	6	1	6	6	1	6
BIRD	1	6	1	6	6	1	6	6	1	6
TREE	1	6	1	6	6	1	6	6	1	6
SEED	1	6	1	6	6	1	6	6	1	6
HAT	1	6	1	6	6	1	6	6	1	6
BED	1	6	1	6	6	1	6	6	1	6
ARM	1	6	1	6	6	1	6	6	1	6
FACE	1	6	1	6	6	1	6	6	1	6
RIVER	1	6	1	6	6	1	6	6	1	6
HOUSE	1	6	1	6	6	1	6	6	1	6
WATER	1	6	1	6	6	1	6	6	1	6
STONE	1	6	1	6	6	1	6	6	1	6
MALE	9	9	9	9	9	9	9	9	9	9
FEMALE	9	9	9	9	9	9	9	9	9	9
WHITE	6	1	6	1	1	6	6	1	6	1
RED	1	1	1	1	1	1	6	1	6	1
ROUND	6	1	6	1	1	6	6	1	6	1
FLAT	6	1	6	1	1	6	6	1	6	1
OLD	6	1	6	1	1	6	6	1	6	1
YOUNG	6	1	6	1	1	6	6	1	6	1
GOOD	6	1	6	1	1	6	6	1	6	1
BAD	6	1	6	1	1	6	6	1	6	1
BIG	6	1	6	1	1	6	1	6	6	1
SHORT	6	1	6	1	1	6	6	1	6	1
HEAVY	6	1	6	1	1	6	6	1	6	1
SOFT	6	1	6	1	1	6	6	1	6	1
HAPPY	6	1	6	1	1	6	6	1	6	1
ANGRY	6	1	6	1	1	6	6	1	6	1
KNOW	6	1	6	1	1	6	6	1	6	1
WANT	6	1	6	1	1	6	6	1	6	1
LOVE	6	1	6	1	1	6	6	1	6	1
SEE/LOOK AT	6	1	6	1	1	6	6	1	6	1
SIT	6	1	6	1	1	6	6	1	6	1
WEAR	6	1	6	1	1	6	6	1	6	1
FIGHT	6	1	6	1	1	6	6	1	6	1
EAT	6	1	6	1	1	6	6	1	6	1
COME	6	1	6	1	1	6	6	1	6	1
COOK	6	1	6	1	1	6	6	1	6	1
BUILD	6	1	6	1	1	6	6	1	6	1
COUGH	6	1	6	1	1	6	6	1	6	1
HIT	6	1	6	1	1	6	6	1	6	1
WAVE	6	1	6	1	1	6	6	1	6	1
DIE	6	1	6	1	1	6	6	1	6	1
GIVE	6	1	6	1	1	6	6	1	6	1
BREAK	6	1	6	1	1	6	6	1	6	1

Appendix B. Parts of speech data coded for MDS analysis (cont'd)

	KH-P-SC-Cop	KH-P-BP-TAM1	KH-P-BP-TAM2	KH-P-BP-0	KS-R-SC-0	KS-R-SC-Aff	KS-R-BP-Class	KS-R-BP-Num	KS-R-BP-0	KS-M-SC-0
WOMAN	1	6	6	1	1	6	1	1	6	1
BOY	1	6	6	1	1	6	1	1	6	9
FATHER	1	6	6	1	1	6	1	1	6	9
SISTER	1	6	1	1	1	6	1	1	6	9
DOG	1	6	6	1	1	6	1	1	6	9
BIRD	1	6	6	1	1	6	1	1	6	9
TREE	1	6	6	1	1	6	1	1	6	9
SEED	1	6	6	1	1	6	1	1	6	9
HAT	1	6	6	1	1	6	1	1	6	9
BED	1	6	6	1	9	9	9	9	9	9
ARM	1	6	6	1	1	6	1	1	6	9
FACE	1	6	6	1	1	6	1	1	6	9
RIVER	1	6	6	1	1	6	1	1	6	9
HOUSE	1	6	6	1	1	6	1	1	6	9
WATER	1	6	6	1	1	6	1	6	6	9
STONE	1	6	6	1	1	6	1	1	6	9
MALE	9	9	9	9	9	9	9	9	9	9
FEMALE	9	9	9	9	9	1	9	9	9	1
WHITE	6	6	1	6	6	1	6	6	1	1
RED	1	6	1	6	6	1	6	6	1	1
ROUND	6	6	1	6	6	1	6	6	1	6
FLAT	6	6	1	6	6	1	6	6	1	1
OLD	6	6	1	6	6	1	6	6	1	1
YOUNG	6	6	1	6	6	1	6	6	1	1
GOOD	6	6	1	6	6	1	6	6	1	1
BAD	6	6	1	6	6	1	6	6	1	6
BIG	6	6	1	6	6	1	6	6	1	1
SHORT	6	6	1	6	6	1	6	6	1	1
HEAVY	6	6	1	6	1	6	9	9	9	9
SOFT	6	6	1	6	6	1	6	6	1	1
HAPPY	6	6	1	6	9	9	9	9	9	9
ANGRY	6	6	1	6	1	6	9	9	9	9
KNOW	6	1	6	6	6	1	6	6	1	6
WANT	6	1	6	6	6	1	6	6	1	6
LOVE	6	1	6	6	1	6	9	9	9	6
SEE/LOOK AT	6	1	6	6	6	1	6	6	1	6
SIT	6	1	6	6	6	1	6	6	1	6
WEAR	6	1	6	6	6	1	6	6	1	6
FIGHT	6	1	6	6	6	1	6	6	1	6
EAT	6	1	6	6	6	1	6	6	1	1
COME	6	1	6	6	6	1	6	6	1	6
COOK	6	1	6	6	6	1	6	6	1	1
BUILD	6	1	6	6	6	1	6	6	1	6
COUGH	6	1	6	6	6	1	6	6	1	6
HIT	6	1	6	6	6	1	6	6	1	6
WAVE	6	1	6	6	6	1	6	6	1	6
DIE	6	1	6	6	6	1	6	6	1	6
GIVE	6	1	6	6	6	1	6	6	1	6
BREAK	6	1	6	6	6	1	6	6	1	6

Appendix B. Parts of speech data coded for MDS analysis (cont'd)

	KS-M-SC-Suff	KS-P-SC-0	KS-P-SC-Cop	KS-P-BP-TAM	KS-P-BP-0	JI-R-SC-0	JI-R-SC-Aff	JI-R-BP-NGC	JI-R-BP-0	JI-M-SC-0
WOMAN	9	6	1	6	1	1	6	1	6	9
BOY	9	6	1	6	1	1	6	1	6	9
FATHER	9	6	1	6	1	1	6	1	6	9
SISTER	9	6	1	6	1	1	6	1	6	9
DOG	9	6	1	6	1	1	6	1	6	9
BIRD	9	6	1	6	1	1	6	1	6	9
TREE	9	6	1	6	1	1	6	1	6	9
SEED	9	6	1	6	1	1	6	1	6	9
HAT	9	6	1	6	1	1	6	1	6	9
BED	9	9	9	9	9	1	6	1	6	9
ARM	9	6	1	6	1	1	6	1	6	9
FACE	9	6	1	6	1	1	6	1	6	9
RIVER	9	6	1	6	1	1	6	1	6	9
HOUSE	9	6	1	6	1	1	6	1	6	9
WATER	9	6	1	6	1	1	6	1	6	9
STONE	9	6	1	6	1	1	6	1	6	9
MALE	9	6	1	6	1	9	9	9	9	1
FEMALE	9	6	1	6	1	9	9	9	9	1
WHITE	1	1	1	1	1	9	9	1	6	1
RED	1	1	1	1	1	9	9	1	6	1
ROUND	1	1	6	1	6	9	9	6	1	9
FLAT	6	6	1	6	1	9	9	6	1	9
OLD	1	1	1	1	1	9	9	6	1	1
YOUNG	6	6	1	6	1	9	9	6	1	1
GOOD	1	1	1	1	1	9	9	6	1	1
BAD	1	1	6	1	6	9	9	6	1	1
BIG	6	6	1	6	1	9	9	6	1	1
SHORT	9	1	1	1	1	9	9	6	1	1
HEAVY	9	1	1	1	1	9	9	6	1	1
SOFT	6	6	1	6	1	9	9	6	1	1
HAPPY	9	9	9	9	9	9	9	6	1	1
ANGRY	9	6	1	6	1	9	9	6	1	1
KNOW	1	1	6	1	6	6	1	6	1	6
WANT	1	1	6	1	6	6	1	6	1	6
LOVE	1	1	1	1	1	6	1	6	1	6
SEE/LOOK AT	1	1	6	1	6	6	1	6	1	6
SIT	1	1	6	1	6	6	1	6	1	6
WEAR	1	1	6	1	6	9	9	6	1	9
FIGHT	1	1	6	1	6	6	1	6	1	6
EAT	9	1	9	1	9	6	1	6	1	6
COME	1	1	6	1	6	6	1	6	1	6
COOK	9	1	9	1	9	6	1	6	1	6
BUILD	1	1	6	1	6	9	9	6	1	9
COUGH	1	1	6	1	6	6	1	6	1	6
HIT	1	1	6	1	6	6	1	6	1	6
WAVE	1	1	6	1	6	6	1	6	1	6
DIE	1	1	6	1	6	6	1	6	1	6
GIVE	1	1	6	1	6	6	1	6	1	6
BREAK	1	1	6	1	6	6	1	6	1	6

Appendix B. Parts of speech data coded for MDS analysis (cont'd)

	JI-M-SC-Aff	JI-M-BP-NGCagr	JI-M-BP-0	JI-P-SC-0	JI-P-SC-Aff	JI-P-SC-LV	JI-P-BP-TAM-abl	JI-P-BP-TAM-peri	JI-P-BP-0	MA-R-BP-NPArt
WOMAN	1	1	6	6	1	6	6	6	1	1
BOY	1	1	6	6	1	6	6	6	1	1
FATHER	1	1	6	6	1	6	6	6	1	1
SISTER	1	1	6	6	1	6	6	6	1	1
DOG	1	1	6	6	1	6	6	6	1	1
BIRD	1	1	6	6	1	6	6	6	1	1
TREE	1	1	6	6	1	6	6	6	1	1
SEED	1	1	6	6	1	6	6	6	1	1
HAT	1	1	6	6	1	6	6	6	1	1
BED	1	1	6	6	1	6	6	6	1	1
ARM	1	1	6	6	1	6	6	6	1	1
FACE	1	1	6	6	1	6	6	6	1	1
RIVER	1	1	6	6	1	6	6	6	1	1
HOUSE	1	1	6	6	1	6	6	6	1	1
WATER	1	1	6	6	1	6	6	6	1	1
STONE	1	1	6	6	1	6	6	6	1	1
MALE	6	1	6	1	6	6	6	6	1	6
FEMALE	6	1	6	1	6	6	6	6	1	6
WHITE	6	1	6	1	6	6	6	6	1	6
RED	6	1	6	1	6	6	6	6	1	1
ROUND	9	9	9	9	9	9	9	9	9	9
FLAT	9	9	9	9	9	9	9	9	9	6
OLD	6	1	6	1	6	6	6	6	1	6
YOUNG	6	1	6	1	6	6	6	6	1	6
GOOD	6	1	6	1	6	6	6	6	1	6
BAD	6	1	6	1	6	6	6	6	1	6
BIG	6	1	6	1	6	6	6	6	1	6
SHORT	6	1	6	1	6	6	6	6	1	6
HEAVY	6	1	6	1	6	6	6	6	1	6
SOFT	6	1	6	1	6	6	6	6	1	6
HAPPY	6	1	6	1	6	6	6	6	1	6
ANGRY	6	1	6	1	6	6	6	6	1	6
KNOW	1	6	1	6	6	1	6	1	6	6
WANT	1	6	1	6	6	1	6	1	6	6
LOVE	1	6	1	6	6	1	6	1	6	6
SEE/LOOK AT	1	6	1	6	6	1	6	1	6	6
SIT	1	6	1	6	6	1	6	1	6	6
WEAR	9	9	9	9	9	9	9	9	9	6
FIGHT	1	6	1	6	6	1	6	1	6	6
EAT	1	6	1	6	6	1	6	1	6	6
COME	1	6	1	1	6	6	1	6	6	6
COOK	1	6	1	6	6	1	6	1	6	6
BUILD	9	9	9	9	9	9	9	9	9	6
COUGH	1	6	1	6	6	1	6	1	6	6
HIT	1	6	1	6	6	1	6	1	6	6
WAVE	1	6	1	6	6	1	6	1	6	6
DIE	1	6	1	6	6	1	6	1	6	6
GIVE	1	6	1	6	6	1	6	1	6	6
BREAK	1	6	1	6	6	1	6	1	6	6

Appendix B. Parts of speech data coded for MDS analysis (cont'd)

	MA-R-BP-SpecArt	MA-R-BP-NClass	MA-R-BP-0	MA-M-SC-REL	MA-M-SC-REL+Pr	MA-M-SC-PossPr	MA-M-BP-Intens	MA-M-BP-0	MA-P-SC-0	MA-P-SC-Cop
WOMAN	1	1	6	6	1	1	6	1	6	1
BOY	1	1	6	6	1	1	6	1	6	1
FATHER	1	1	6	6	1	1	6	1	6	1
SISTER	1	1	6	6	1	1	6	1	6	1
DOG	1	1	6	6	1	1	6	1	6	1
BIRD	1	1	6	6	1	1	6	1	6	1
TREE	1	1	6	6	1	1	6	1	6	1
SEED	1	1	6	6	1	1	6	1	6	1
HAT	1	1	6	6	1	1	6	1	6	1
BED	1	1	6	6	1	1	6	1	6	1
ARM	1	1	6	6	1	1	6	1	6	1
FACE	1	1	6	6	1	1	6	1	6	1
RIVER	1	1	6	6	1	1	6	1	6	1
HOUSE	1	1	6	6	1	1	6	1	6	1
WATER	1	1	6	6	1	1	6	1	6	1
STONE	1	1	6	6	1	1	6	1	6	1
MALE	1	1	6	1	6	6	1	6	6	1
FEMALE	1	1	6	1	6	6	1	6	6	1
WHITE	1	1	6	1	6	6	1	6	6	1
RED	1	1	6	6	1	1	6	1	6	1
ROUND	9	9	9	9	9	9	9	9	9	9
FLAT	1	1	6	1	6	6	1	6	6	1
OLD	1	1	6	1	6	6	1	6	6	1
YOUNG	1	1	6	1	6	6	1	6	6	1
GOOD	1	1	6	1	6	6	1	6	6	1
BAD	1	1	6	1	6	6	1	6	6	1
BIG	1	1	6	1	6	6	1	6	6	1
SHORT	1	1	6	1	6	6	1	6	6	1
HEAVY	1	1	6	1	6	6	1	6	6	1
SOFT	1	1	6	1	6	6	1	6	6	1
HAPPY	6	6	1	1	6	6	6	1	1	6
ANGRY	6	6	1	1	6	6	6	1	1	6
KNOW	6	6	1	1	6	6	6	1	1	6
WANT	6	6	1	1	6	6	6	1	1	6
LOVE	6	6	1	1	6	6	6	1	1	6
SEE/LOOK AT	6	6	1	1	6	6	6	1	1	6
SIT	6	6	1	1	6	6	6	1	1	6
WEAR	6	6	1	1	6	6	6	1	1	6
FIGHT	6	6	1	1	6	6	6	1	1	6
EAT	6	6	1	1	6	6	6	1	1	6
COME	6	6	1	1	6	6	6	1	1	6
COOK	6	6	1	1	6	6	6	1	1	6
BUILD	6	6	1	1	6	6	6	1	1	6
COUGH	6	6	1	1	6	6	6	1	1	6
HIT	6	6	1	1	6	6	6	1	1	6
WAVE	6	6	1	1	6	6	6	1	1	6
DIE	6	6	1	1	6	6	6	1	1	6
GIVE	6	6	1	1	6	6	6	1	1	6
BREAK	6	6	1	1	6	6	6	1	1	6

Appendix B. Parts of speech data coded for MDS analysis (cont'd)

	MA-P-BP-ConcPr	MA-P-BP-T1	MA-P-BP-T2	MA-P-BP-T3	MA-P-BP-T4	TO-R-SC-0	TO-R-SC-Aff	TO-R-BP-Num	TO-R-BP-Pron	TO-R-BP-PLmark
WOMAN	6	6	6	6	1	1	6	1	1	1
BOY	6	6	6	6	1	1	6	1	1	1
FATHER	6	6	6	6	1	1	6	1	1	1
SISTER	6	6	6	6	1	1	6	1	1	1
DOG	6	6	6	6	1	1	6	1	6	1
BIRD	6	6	6	6	1	1	6	1	6	6
TREE	6	6	6	6	1	1	6	1	6	6
SEED	6	6	6	6	1	1	6	1	6	6
HAT	6	6	6	6	1	1	6	1	6	6
BED	6	6	6	6	1	1	6	1	6	6
ARM	6	6	6	6	1	1	6	1	6	6
FACE	6	6	6	6	1	1	6	1	6	6
RIVER	6	6	6	6	1	9	9	9	9	9
HOUSE	6	6	6	6	1	1	6	1	6	6
WATER	6	6	6	6	1	1	6	1	6	6
STONE	6	6	6	6	1	1	6	1	6	6
MALE	1	6	6	6	1	9	9	9	9	9
FEMALE	1	6	6	6	1	9	9	9	9	9
WHITE	1	6	6	6	1	6	1	6	6	6
RED	6	6	6	6	1	6	1	6	6	6
ROUND	9	9	9	9	9	6	1	6	6	6
FLAT	1	6	6	6	1	6	1	6	6	6
OLD	1	6	6	6	1	6	1	6	6	6
YOUNG	1	6	6	6	1	6	1	6	6	6
GOOD	1	6	6	6	1	6	1	6	6	6
BAD	1	6	6	6	1	6	1	6	6	6
BIG	1	6	6	6	1	6	1	6	6	6
SHORT	1	6	6	6	1	6	1	6	6	6
HEAVY	1	6	6	6	1	6	1	6	6	6
SOFT	1	6	6	6	1	6	1	6	6	6
HAPPY	1	6	6	6	1	6	1	6	6	6
ANGRY	1	6	6	6	1	6	1	6	6	6
KNOW	1	6	1	6	6	6	1	6	6	6
WANT	1	6	1	6	6	6	1	6	6	6
LOVE	1	6	1	6	6	6	1	6	6	6
SEE/LOOK AT	1	6	1	6	6	6	1	6	6	6
SIT	1	1	6	6	6	6	1	6	6	6
WEAR	1	1	6	6	6	6	1	6	6	6
FIGHT	1	6	6	1	6	6	1	6	6	6
EAT	1	1	6	6	6	6	1	6	6	6
COME	1	1	6	6	6	6	1	6	6	6
COOK	1	6	1	6	6	6	1	6	6	6
BUILD	1	6	1	6	6	6	1	6	6	6
COUGH	1	6	1	6	6	6	1	6	6	6
HIT	1	6	6	1	6	6	1	6	6	6
WAVE	1	6	1	6	6	9	9	9	9	9
DIE	1	6	1	6	6	6	1	6	6	6
GIVE	1	6	1	6	6	6	1	6	6	6
BREAK	1	6	1	6	6	6	1	6	6	6

Appendix B. Parts of speech data coded for MDS analysis (cont'd)

	TO-R-BP-Subj-IndNum	TO-R-BP-0	TO-M-SC-0	TO-M-SC-Aff	TO-M-SC-RelMar	TO-P-BP-TAM	TO-P-BP-0	QE-R-SC-0	QE-R-SC-Aff	QE-M-SC-0
WOMAN	1	6	1	1	6	6	1	1	6	1
BOY	1	6	1	1	6	6	1	1	6	1
FATHER	1	6	1	1	6	6	1	1	6	1
SISTER	1	6	1	1	6	6	1	1	6	1
DOG	1	6	1	1	6	6	1	1	6	1
BIRD	6	6	1	1	6	6	1	1	6	1
TREE	6	6	1	1	6	6	1	1	6	1
SEED	6	6	1	1	6	6	1	1	6	1
HAT	6	6	1	1	6	6	1	1	6	1
BED	6	6	1	1	6	6	1	1	6	1
ARM	6	6	1	1	6	6	1	1	6	1
FACE	6	6	1	1	6	6	1	1	6	1
RIVER	9	9	9	9	9	9	9	1	6	1
HOUSE	6	6	1	1	6	6	1	1	6	1
WATER	6	6	1	1	6	6	1	1	6	1
STONE	6	6	1	1	6	6	1	1	6	1
MALE	9	9	1	1	6	6	1	9	9	1
FEMALE	9	9	1	1	6	6	1	9	9	1
WHITE	6	1	1	6	1	1	6	9	9	1
RED	6	1	1	6	1	1	6	9	9	1
ROUND	6	1	1	6	1	1	6	9	9	1
FLAT	6	1	1	6	1	1	6	9	9	1
OLD	6	1	1	6	1	1	6	9	9	1
YOUNG	6	1	1	6	1	1	6	9	9	1
GOOD	6	1	1	6	1	1	6	9	9	1
BAD	6	1	1	6	1	1	6	9	9	1
BIG	6	1	1	6	1	1	6	9	9	1
SHORT	6	1	1	6	1	1	6	9	9	1
HEAVY	6	1	1	6	1	1	6	9	1	1
SOFT	6	1	1	6	1	1	6	9	1	1
HAPPY	6	1	1	6	1	1	6	6	1	6
ANGRY	6	1	1	6	1	1	6	9	1	1
KNOW	6	1	6	6	1	1	6	6	1	6
WANT	6	1	6	6	1	1	6	6	1	6
LOVE	6	1	6	6	1	1	6	6	1	6
SEE/LOOK AT	6	1	6	6	1	1	6	6	1	6
SIT	6	1	6	6	1	1	6	6	1	6
WEAR	6	1	6	6	1	1	6	6	1	6
FIGHT	6	1	6	6	1	1	6	6	1	6
EAT	6	1	6	6	1	1	6	6	1	6
COME	6	1	6	6	1	1	6	6	1	6
COOK	6	1	6	6	1	1	6	6	1	6
BUILD	6	1	6	6	1	1	6	6	1	6
COUGH	6	1	6	6	1	1	6	6	1	6
HIT	6	1	6	6	1	1	6	6	1	6
WAVE	9	9	9	9	9	9	9	6	1	6
DIE	6	1	6	6	1	1	6	6	1	6
GIVE	6	1	6	6	1	1	6	6	1	6
BREAK	6	1	6	6	1	1	6	6	1	6

Appendix B. Parts of speech data coded for MDS analysis (cont'd)

	QE-M-SC-Suff	QE-M-BP-SPRL	QE-M-BP-0	QE-P-SC-0	QE-P-SC-Cop	QE-P-BP-TAM	QE-P-BP-TAM-peri	CR-R-SC-0	CR-R-SC-Aff	CR-R-BP-Num
WOMAN	1	6	1	6	1	6	1	1	6	1
BOY	1	6	1	6	1	6	1	1	6	1
FATHER	1	6	1	6	1	6	1	1	6	6
SISTER	1	6	1	6	1	6	1	1	6	6
DOG	1	6	1	6	1	6	1	1	6	6
BIRD	1	6	1	6	1	6	1	1	6	6
TREE	1	6	1	6	1	6	1	1	6	6
SEED	1	6	1	6	1	6	1	1	6	6
HAT	1	6	1	6	1	6	1	1	6	6
BED	1	6	1	6	1	6	1	1	6	6
ARM	1	6	1	6	1	6	1	1	6	6
FACE	1	6	1	6	1	6	1	1	6	6
RIVER	1	6	1	6	1	6	1	1	6	6
HOUSE	1	6	1	6	1	6	1	1	6	6
WATER	1	6	1	6	1	6	1	1	6	6
STONE	1	1	6	6	1	6	1	1	6	6
MALE	6	1	6	6	1	6	1	9	9	9
FEMALE	6	1	6	6	1	6	1	9	9	9
WHITE	6	1	6	6	1	6	1	6	1	6
RED	6	1	6	6	1	6	1	6	1	6
ROUND	6	1	6	6	1	6	1	6	1	6
FLAT	6	1	6	6	1	6	1	6	1	6
OLD	6	1	6	6	1	6	1	6	1	6
YOUNG	6	1	6	6	1	6	1	6	1	6
GOOD	6	1	6	6	1	6	1	6	1	6
BAD	6	1	6	6	1	6	1	6	1	6
BIG	6	1	6	6	1	6	1	6	1	6
SHORT	6	1	6	6	1	6	1	6	1	6
HEAVY	1	1	1	1	1	1	1	6	1	6
SOFT	1	1	1	1	1	1	1	6	1	6
HAPPY	1	6	1	1	6	1	6	6	1	6
ANGRY	1	1	1	1	1	1	1	6	1	6
KNOW	1	6	1	1	6	1	6	6	1	6
WANT	1	6	1	1	6	1	6	6	1	6
LOVE	1	6	1	1	6	1	6	6	1	6
SEE/LOOK AT	1	6	1	1	6	1	6	6	1	6
SIT	1	6	1	1	6	1	6	6	1	6
WEAR	1	6	1	1	6	1	6	6	1	6
FIGHT	1	6	1	1	6	1	6	6	1	6
EAT	1	6	1	1	6	1	6	6	1	6
COME	1	6	1	1	6	1	6	6	1	6
COOK	1	6	1	1	6	1	6	6	1	6
BUILD	1	6	1	1	6	1	6	6	1	6
COUGH	1	6	1	1	6	1	6	6	1	6
HIT	1	6	1	1	6	1	6	6	1	6
WAVE	1	6	1	1	6	1	6	6	1	6
DIE	1	6	1	1	6	1	6	6	1	6
GIVE	1	6	1	1	6	1	6	6	1	6
BREAK	1	6	1	1	6	1	6	6	1	6

Appendix B. Parts of speech data coded for MDS analysis (cont'd)

	CR-R-BP-Dim	CR-R-BP-0	CR-M-SC-Attach	CR-M-SC-Aff	CR-M-SC-AffLgr	CR-M-BP-CMPR	CR-M-BP-0	CR-P-SC-0	CR-P-SC-DURAff	CR-P-SC-Cop
WOMAN	1	6	6	1	6	6	1	6	6	1
BOY	1	6	6	1	6	6	1	6	6	1
FATHER	1	6	6	1	6	6	1	6	6	1
SISTER	1	6	6	1	6	6	1	6	6	1
DOG	1	6	6	1	6	6	1	6	6	1
BIRD	1	6	6	1	6	6	1	6	6	1
TREE	1	6	6	1	6	6	1	6	6	1
SEED	1	6	6	1	6	6	1	6	6	1
HAT	1	6	6	1	6	6	1	6	6	1
BED	1	6	6	1	6	6	1	6	6	1
ARM	1	6	6	1	6	6	1	6	6	1
FACE	1	6	6	1	6	6	1	6	6	1
RIVER	1	6	6	1	6	6	1	6	6	1
HOUSE	1	6	6	1	6	6	1	6	6	1
WATER	1	6	6	1	6	6	1	6	6	1
STONE	1	6	6	1	6	6	1	6	6	1
MALE	9	9	9	9	9	9	9	9	9	9
FEMALE	9	9	9	9	9	9	9	9	9	9
WHITE	6	1	1	1	6	1	6	6	1	1
RED	6	1	1	1	6	1	6	6	1	1
ROUND	6	1	1	1	6	1	6	6	1	1
FLAT	6	1	1	1	6	1	6	6	1	1
OLD	6	1	1	1	6	1	6	6	1	1
YOUNG	6	1	1	1	6	1	6	6	1	1
GOOD	6	1	1	1	6	1	6	6	1	1
BAD	6	1	1	1	6	1	6	6	1	1
BIG	6	1	1	1	6	1	6	6	1	1
SHORT	6	1	1	1	6	1	6	6	1	1
HEAVY	6	1	1	1	6	1	6	6	1	1
SOFT	6	1	1	1	6	1	6	6	1	1
HAPPY	6	1	1	1	6	1	6	6	1	1
ANGRY	6	1	1	1	6	1	6	6	1	1
KNOW	6	1	6	1	6	1	6	1	6	6
WANT	6	1	6	1	6	1	6	1	6	6
LOVE	6	1	6	1	6	1	6	1	6	6
SEE/LOOK AT	6	1	6	1	6	1	6	1	6	6
SIT	6	1	6	1	6	1	6	1	6	6
WEAR	6	1	6	1	6	1	6	1	6	6
FIGHT	6	1	6	6	1	6	1	1	6	6
EAT	6	1	6	6	1	6	1	1	6	6
COME	6	1	6	6	1	6	1	1	6	6
COOK	6	1	6	6	1	6	1	1	6	6
BUILD	6	1	6	6	1	6	1	1	6	6
COUGH	6	1	6	6	1	6	1	1	6	6
HIT	6	1	6	6	1	6	1	1	6	6
WAVE	6	1	6	6	1	6	1	1	6	6
DIE	6	1	6	6	1	6	1	1	6	6
GIVE	6	1	6	6	1	6	1	1	6	6
BREAK	6	1	6	6	1	6	1	1	6	6

Appendix B. Parts of speech data coded for MDS analysis (cont'd)

	CR-P-BP-TAM	CR-P-BP-TAM-peri	KI-R-SC-0	KI-R-SC-Aff	KI-R-BP-Num-Suff	KI-R-BP-Num-Peri	KI-R-BP-Num-Vindex	KI-R-BP-0	KI-M-SC-0	KI-M-SC-GEN
WOMAN	6	1	1	6	1	6	1	6	6	1
BOY	6	1	1	6	1	6	1	6	6	1
FATHER	6	1	1	6	6	1	1	6	6	1
SISTER	6	1	1	6	6	1	1	6	6	1
DOG	6	1	1	6	6	1	1	6	6	1
BIRD	6	1	1	6	6	1	1	6	6	1
TREE	6	1	1	6	6	1	6	6	6	1
SEED	6	1	1	6	6	1	6	6	6	1
HAT	6	1	1	6	6	1	6	6	6	1
BED	6	1	1	6	6	1	6	6	6	1
ARM	6	1	1	6	6	1	6	6	6	1
FACE	6	1	1	6	6	1	6	6	6	1
RIVER	6	1	1	6	6	1	6	6	6	1
HOUSE	6	1	1	6	6	1	6	6	6	1
WATER	6	1	1	6	6	1	6	6	6	1
STONE	6	1	1	6	6	1	6	6	6	1
MALE	9	9	9	9	9	9	9	9	1	6
FEMALE	9	9	9	9	9	9	9	9	1	6
WHITE	1	1	1	6	6	6	6	1	1	6
RED	1	1	1	6	6	6	6	1	1	6
ROUND	1	1	9	9	9	9	9	9	1	6
FLAT	1	1	9	9	9	9	9	9	1	6
OLD	1	1	6	1	6	6	6	1	1	6
YOUNG	1	1	6	1	6	6	6	1	1	6
GOOD	1	1	6	1	6	6	6	1	1	6
BAD	1	1	6	1	6	6	6	1	1	6
BIG	1	1	6	1	6	6	6	1	1	6
SHORT	1	1	6	1	6	6	6	1	1	6
HEAVY	1	1	6	1	6	6	6	1	1	6
SOFT	1	1	6	1	6	6	6	1	1	6
HAPPY	1	1	6	1	6	6	6	1	6	6
ANGRY	1	1	6	1	6	6	6	1	1	6
KNOW	1	6	6	1	6	6	6	1	6	6
WANT	1	6	6	1	6	6	6	1	6	6
LOVE	1	6	6	1	6	6	6	1	6	6
SEE/LOOK AT	1	6	6	1	6	6	6	1	6	6
SIT	1	6	6	1	6	6	6	1	6	6
WEAR	1	6	6	1	6	6	6	1	6	6
FIGHT	1	6	6	1	6	6	6	1	6	6
EAT	1	6	6	1	6	6	6	1	6	6
COME	1	6	6	1	6	6	6	1	6	6
COOK	1	6	6	1	6	6	6	1	6	6
BUILD	1	6	6	1	6	6	6	1	6	6
COUGH	1	6	6	1	6	6	6	1	6	6
HIT	1	6	6	1	6	6	6	1	6	6
WAVE	1	6	9	9	9	9	9	9	9	9
DIE	1	6	6	1	6	6	6	1	6	6
GIVE	1	6	6	1	6	6	6	1	6	6
BREAK	1	6	6	1	6	6	6	1	6	6

Appendix B. Parts of speech data coded for MDS analysis (cont'd)

	KI-M-SC-REL	KI-M-BP-CMPR	KI-M-BP-0	KI-P-SC-0	KI-P-SC-Plus-Noun	KI-P-SC-FOC	KI-P-BP-TAM-IndArg	KI-P-BP-0	EN-R-SC-0	EN-R-SC-Aff
WOMAN	6	6	1	6	6	1	6	1	1	6
BOY	6	6	1	6	6	1	6	1	1	6
FATHER	6	6	1	6	6	1	6	1	1	6
SISTER	6	6	1	6	6	1	6	1	1	6
DOG	6	6	1	6	6	1	6	1	1	6
BIRD	6	6	1	6	6	1	6	1	1	6
TREE	6	6	1	6	6	1	6	1	1	6
SEED	6	6	1	6	6	1	6	1	1	6
HAT	6	6	1	6	6	1	6	1	1	6
BED	6	6	1	6	6	1	6	1	1	6
ARM	6	6	1	6	6	1	6	1	1	6
FACE	6	6	1	6	6	1	6	1	1	6
RIVER	6	6	1	6	6	1	6	1	1	6
HOUSE	6	6	1	6	6	1	6	1	1	6
WATER	6	6	1	6	6	1	6	1	1	6
STONE	6	6	1	1	6	6	6	1	1	6
MALE	6	6	1	1	6	6	6	1	6	1
FEMALE	6	6	1	1	6	6	6	1	6	1
WHITE	6	1	6	1	6	6	6	1	1	1
RED	6	1	6	1	6	6	6	1	1	1
ROUND	6	1	6	1	6	6	6	1	6	1
FLAT	6	1	6	1	6	6	6	1	6	1
OLD	6	1	6	1	6	6	6	1	6	6
YOUNG	6	1	6	6	1	6	6	1	6	1
GOOD	6	1	6	1	6	6	6	1	6	1
BAD	6	1	6	1	6	6	6	1	6	6
BIG	6	1	6	1	6	6	6	1	6	1
SHORT	6	1	6	1	6	6	6	1	6	1
HEAVY	6	1	6	1	6	6	6	1	6	1
SOFT	6	1	6	1	6	6	6	1	6	1
HAPPY	1	1	6	1	6	6	1	6	6	1
ANGRY	6	1	6	1	6	6	6	1	1	6
KNOW	1	6	1	1	6	6	1	6	6	1
WANT	1	6	1	1	6	6	1	6	6	1
LOVE	1	6	1	1	6	6	1	6	6	1
SEE/LOOK AT	1	6	1	1	6	6	1	6	6	1
SIT	1	6	1	1	6	6	1	6	6	1
WEAR	1	6	1	1	6	6	1	6	6	1
FIGHT	1	6	1	1	6	6	1	6	6	1
EAT	1	6	1	1	6	6	1	6	6	1
COME	1	6	1	1	6	6	1	6	6	1
COOK	1	6	1	1	6	6	1	6	6	1
BUILD	1	6	1	1	6	6	1	6	6	1
COUGH	1	6	1	1	6	6	1	6	6	1
HIT	1	6	1	1	6	6	1	6	6	1
WAVE	9	9	1	9	9	9	9	9	6	1
DIE	1	6	1	1	6	6	1	6	6	1
GIVE	1	6	1	1	6	6	1	6	6	1
BREAK	1	6	1	1	6	6	1	6	6	1

Appendix B. Parts of speech data coded for MDS analysis (cont'd)

	EN-R-SC-Peri	EN-R-BP-Num-Abl	EN-R-BP-Num-Suff	EN-R-BP-0	EN-M-SC-0	EN-M-SC-Aff	EN-M-SC-REL	EN-M-BP-CMPR-Abl	EN-M-BP-CMPR-Suff	EN-M-BP-CMPR-Suff2
WOMAN	6	1	6	6	1	1	6	6	6	6
BOY	6	6	1	6	1	1	6	6	6	6
FATHER	6	6	1	6	6	1	6	6	6	6
SISTER	6	6	1	6	6	1	6	6	6	6
DOG	6	6	1	6	1	1	6	6	6	6
BIRD	6	6	1	6	1	1	6	6	6	6
TREE	6	6	1	6	1	1	6	6	6	6
SEED	6	6	1	6	1	1	6	6	6	6
HAT	6	6	1	6	1	1	6	6	6	6
BED	6	6	1	6	1	1	6	6	6	6
ARM	6	6	1	6	1	1	6	6	6	6
FACE	6	6	1	6	1	1	6	6	6	6
RIVER	6	6	1	6	1	1	6	6	6	6
HOUSE	6	6	1	6	1	1	6	6	6	6
WATER	6	6	1	6	1	1	6	6	6	6
STONE	6	6	1	6	1	1	6	6	6	6
MALE	6	6	6	1	1	6	6	6	6	6
FEMALE	6	6	6	1	1	6	6	6	6	6
WHITE	6	6	6	1	1	1	6	6	1	6
RED	6	6	6	1	1	1	6	6	1	6
ROUND	6	6	6	1	1	6	6	6	6	6
FLAT	6	6	6	1	1	6	6	6	1	6
OLD	1	6	6	1	1	6	6	6	1	6
YOUNG	6	6	6	1	1	6	6	6	1	6
GOOD	6	6	6	1	1	6	6	1	6	6
BAD	1	6	6	1	1	6	6	1	6	6
BIG	6	6	6	1	1	6	6	6	1	6
SHORT	6	6	6	1	1	6	6	6	1	6
HEAVY	6	6	6	1	1	6	6	6	1	6
SOFT	6	6	6	1	1	6	6	6	1	6
HAPPY	6	6	6	1	1	6	6	6	1	6
ANGRY	6	6	6	1	6	1	6	6	6	1
KNOW	6	6	6	1	6	1	1	6	6	6
WANT	6	6	6	1	6	1	1	6	6	6
LOVE	6	6	6	1	6	1	1	6	6	6
SEE/LOOK AT	6	6	6	1	6	1	1	6	6	6
SIT	6	6	6	1	6	1	1	6	6	6
WEAR	6	6	6	1	6	1	1	6	6	6
FIGHT	6	6	6	1	6	1	1	6	6	6
EAT	6	6	6	1	6	1	1	6	6	6
COME	6	6	6	1	6	1	1	6	6	6
COOK	6	6	6	1	6	1	1	6	6	6
BUILD	6	6	6	1	6	1	1	6	6	6
COUGH	6	6	6	1	6	1	1	6	6	6
HIT	6	6	6	1	6	1	1	6	6	6
WAVE	6	6	6	1	6	1	1	6	6	6
DIE	6	6	6	1	6	1	1	6	6	6
GIVE	6	6	6	1	6	1	1	6	6	6
BREAK	6	6	6	1	6	1	1	6	6	6

Appendix B. Parts of speech data coded for MDS analysis (cont'd)

	EN-M-BP-CMPR-Peri	EN-M-BP-SPRL-Abl	EN-M-BP-SPRL-Suff	EN-M-BP-SPRL-Suff2	EN-M-BP-0	EN-P-SC-0	EN-P-SC-Cop	EN-P-SC-CopArt	EN-P-BP-TAbl	EN-P-BP-Taff
WOMAN	6	6	6	6	1	6	6	1	6	6
BOY	6	6	6	6	1	6	6	1	6	6
FATHER	6	6	6	6	1	6	6	1	6	6
SISTER	6	6	6	6	1	6	6	1	6	6
DOG	6	6	6	6	1	6	6	1	6	6
BIRD	6	6	6	6	1	6	6	1	6	6
TREE	6	6	6	6	1	6	6	1	6	6
SEED	6	6	6	6	1	6	6	1	6	6
HAT	6	6	6	6	1	6	6	1	6	6
BED	6	6	6	6	1	6	6	1	6	6
ARM	6	6	6	6	1	6	6	1	6	6
FACE	6	6	6	6	1	6	6	1	6	6
RIVER	6	6	6	6	1	6	6	1	6	6
HOUSE	6	6	6	6	1	6	6	1	6	6
WATER	6	6	6	6	1	6	1	6	6	6
STONE	6	6	6	6	1	6	1	6	6	6
MALE	6	6	6	6	1	6	1	6	6	6
FEMALE	6	6	6	6	1	6	1	6	6	6
WHITE	6	6	1	6	6	6	1	6	6	6
RED	6	6	1	6	6	6	1	6	6	6
ROUND	1	6	1	6	6	6	1	6	6	6
FLAT	6	6	1	6	6	6	1	6	6	6
OLD	6	6	1	6	6	6	1	6	6	6
YOUNG	6	6	1	6	6	6	1	6	6	6
GOOD	6	1	6	6	6	6	1	6	6	6
BAD	6	1	6	6	6	6	1	6	6	6
BIG	6	6	1	6	6	6	1	6	6	6
SHORT	6	6	1	6	6	6	1	6	6	6
HEAVY	6	6	1	6	6	6	1	6	6	6
SOFT	6	6	1	6	6	6	1	6	6	6
HAPPY	6	6	1	6	6	6	1	6	6	6
ANGRY	6	6	6	1	6	6	1	6	6	6
KNOW	6	6	6	6	1	1	6	6	1	6
WANT	6	6	6	6	1	1	6	6	6	1
LOVE	6	6	6	6	1	1	6	6	6	1
SEE/LOOK AT	6	6	6	6	1	1	6	6	1	6
SIT	6	6	6	6	1	1	6	6	1	6
WEAR	6	6	6	6	1	1	6	6	1	6
FIGHT	6	6	6	6	1	1	6	6	1	6
EAT	6	6	6	6	1	1	6	6	1	6
COME	6	6	6	6	1	1	6	6	1	6
COOK	6	6	6	6	1	1	6	6	6	1
BUILD	6	6	6	6	1	1	6	6	1	6
COUGH	6	6	6	6	1	1	6	6	6	1
HIT	6	6	6	6	1	1	6	6	1	6
WAVE	6	6	6	6	1	1	6	6	6	1
DIE	6	6	6	6	1	1	6	6	6	1
GIVE	6	6	6	6	1	1	6	6	1	6
BREAK	6	6	6	6	1	1	6	6	1	6

Appendix B. Parts of speech data coded for MDS analysis (cont'd)

	EN-P-BP-TPeri	EN-P-BP-P+N-Suff	EN-P-BP-P+M-Peri	IN-R-SC-0	IN-R-SC-Aff	IN-R-BP-Case-Num	IN-R-BP-0	IN-M-SC-0	IN-M-SC-Aff	IN-M-BP-CMPR
WOMAN	1	6	1	1	6	1	6	6	1	6
BOY	1	6	1	1	6	1	6	6	1	6
FATHER	1	6	1	1	6	1	6	6	1	6
SISTER	1	6	1	1	6	1	6	6	1	6
DOG	1	6	1	1	6	1	6	6	1	6
BIRD	1	6	1	1	6	1	6	6	1	6
TREE	1	6	1	1	6	1	6	6	1	6
SEED	1	6	1	1	6	1	6	6	1	6
HAT	1	6	1	1	6	1	6	6	1	6
BED	1	6	1	1	6	1	6	6	1	6
ARM	1	6	1	1	6	1	6	6	1	6
FACE	1	6	1	1	6	1	6	6	1	6
RIVER	1	6	1	1	6	1	6	6	1	6
HOUSE	1	6	1	1	6	1	6	6	1	6
WATER	1	6	1	1	6	1	6	6	1	6
STONE	1	6	1	1	6	1	6	6	1	6
MALE	1	6	1	6	1	1	6	1	6	6
FEMALE	1	6	1	6	1	1	6	1	6	6
WHITE	1	6	1	6	1	1	6	1	6	1
RED	1	6	1	6	1	1	6	1	6	1
ROUND	1	6	1	6	1	1	6	1	6	1
FLAT	1	6	1	6	1	1	6	1	6	1
OLD	1	6	1	6	1	1	6	1	6	1
YOUNG	1	6	1	6	1	1	6	1	6	1
GOOD	1	6	1	6	1	1	6	1	6	1
BAD	1	6	1	1	1	1	6	1	1	1
BIG	1	6	1	6	1	1	6	1	6	1
SHORT	1	6	1	6	1	1	6	1	6	1
HEAVY	1	6	1	6	1	1	6	1	6	1
SOFT	1	6	1	6	1	1	6	1	6	1
HAPPY	1	6	1	1	1	1	6	1	1	1
ANGRY	1	6	1	6	1	1	6	1	6	1
KNOW	6	1	6	6	1	6	1	6	1	6
WANT	6	1	6	6	1	6	1	6	1	6
LOVE	6	1	6	6	1	6	1	6	1	6
SEE/LOOK AT	6	1	6	6	1	6	1	6	1	6
SIT	6	1	6	6	1	6	1	6	1	6
WEAR	6	1	6	6	1	6	1	6	1	6
FIGHT	6	1	6	6	1	6	1	6	1	6
EAT	6	1	6	6	1	6	1	6	1	6
COME	6	1	6	6	1	6	1	6	1	6
COOK	6	1	6	6	1	6	1	6	1	6
BUILD	6	1	6	6	1	6	1	6	1	6
COUGH	6	1	6	6	1	6	1	6	1	6
HIT	6	1	6	6	1	6	1	6	1	6
WAVE	6	1	6	6	1	6	1	6	1	6
DIE	6	1	6	6	1	6	1	6	1	6
GIVE	6	1	6	6	1	6	1	6	1	6
BREAK	6	1	6	6	1	6	1	6	1	6

Appendix B. Parts of speech data coded for MDS analysis (cont'd)

	IN-M-BP-Gend-Agr	IN-M-BP-Num-Agr	IN-M-BP-Case-Agr	IN-M-BP-0	IN-P-SC-0	IN-P-SC-Comp	IN-P-SC-Cop	IN-P-BP-Iter	IN-P-BP-Aux	IN-P-BP-Ind-Num
WOMAN	6	6	6	1	6	6	1	6	6	6
BOY	6	6	6	1	6	6	1	6	6	6
FATHER	6	6	6	1	6	6	1	6	6	6
SISTER	6	6	6	1	6	6	1	6	6	6
DOG	6	6	6	1	6	6	1	6	6	6
BIRD	6	6	6	1	6	6	1	6	6	6
TREE	6	6	6	1	6	6	1	6	6	6
SEED	6	6	6	1	6	6	1	6	6	6
HAT	6	6	6	1	6	6	1	6	6	6
BED	6	6	6	1	6	6	1	6	6	6
ARM	6	6	6	1	6	6	1	6	6	6
FACE	6	6	6	1	6	6	1	6	6	6
RIVER	6	6	6	1	6	6	1	6	6	6
HOUSE	6	6	6	1	6	6	1	6	6	6
WATER	6	6	6	1	6	6	1	6	6	6
STONE	6	6	6	1	6	6	1	6	6	6
MALE	6	6	1	6	6	6	1	6	6	6
FEMALE	6	6	1	6	6	6	1	6	6	6
WHITE	6	6	1	6	6	6	1	6	6	6
RED	6	6	1	6	6	6	1	6	6	6
ROUND	6	6	1	6	6	6	1	6	6	6
FLAT	6	6	1	6	6	6	1	6	6	6
OLD	6	6	1	6	6	6	1	6	6	6
YOUNG	6	6	1	6	6	6	1	6	6	6
GOOD	6	6	1	6	6	6	1	6	6	6
BAD	6	6	1	6	6	6	1	6	6	6
BIG	1	1	1	6	6	6	1	6	6	6
SHORT	6	6	1	6	6	6	1	6	6	6
HEAVY	1	6	1	6	6	6	1	6	6	6
SOFT	6	6	1	6	6	6	1	6	6	6
HAPPY	6	6	1	6	6	6	1	6	6	6
ANGRY	6	6	1	6	6	6	1	6	6	6
KNOW	6	6	1	6	1	6	6	6	6	6
WANT	6	6	1	6	1	6	6	6	6	6
LOVE	6	6	1	6	1	6	6	6	6	6
SEE/LOOK AT	6	6	1	6	1	6	6	6	6	6
SIT	6	6	1	6	1	6	6	6	6	6
WEAR	6	6	1	6	1	6	6	6	6	6
FIGHT	6	6	1	6	1	6	6	6	6	6
EAT	6	6	1	6	1	6	6	6	6	6
COME	6	6	1	6	1	6	6	6	1	1
COOK	6	6	1	6	6	1	6	6	6	6
BUILD	6	6	1	6	1	6	6	6	1	6
COUGH	6	6	1	6	6	1	6	6	6	6
HIT	6	6	1	6	1	6	6	1	1	6
WAVE	6	6	1	6	1	6	6	6	6	6
DIE	6	6	1	6	1	6	6	6	6	1
GIVE	6	6	1	6	1	6	6	1	1	6
BREAK	6	6	1	6	1	6	6	6	6	6

Appendix B. Parts of speech data coded for MDS analysis (cont'd)

	IN-P-BP-Ind-Gend	IN-P-BP-Sppl	IN-P-BP-Full-Akt	IN-P-BP-Part-Akt	IN-P-BP-Restr-Akt	KA-R-SC-0	KA-R-SC-Aff	KA-R-BP-PL	KA-R-BP-0	KA-M-SC-0
WOMAN	6	6	6	6	1	1	6	1	6	1
BOY	6	6	6	6	1	1	6	1	6	1
FATHER	6	6	6	6	1	1	6	1	6	1
SISTER	6	6	6	6	1	1	6	1	6	1
DOG	6	6	6	6	1	1	6	1	6	1
BIRD	6	6	6	6	1	1	6	1	6	1
TREE	6	6	6	6	1	1	6	1	6	1
SEED	6	6	6	6	1	1	6	1	6	1
HAT	6	6	6	6	1	1	6	1	6	1
BED	6	6	6	6	1	1	6	1	6	1
ARM	6	6	6	6	1	1	6	1	6	1
FACE	6	6	6	6	1	1	6	1	6	1
RIVER	6	6	6	6	1	1	6	1	6	1
HOUSE	6	6	6	6	1	1	6	1	6	1
WATER	6	6	6	6	1	1	6	1	6	1
STONE	6	6	6	6	1	1	6	1	6	1
MALE	6	6	6	6	1	9	1	1	6	1
FEMALE	6	6	6	6	1	9	1	1	6	1
WHITE	6	6	6	6	1	9	1	1	6	1
RED	6	6	6	6	1	1	1	1	6	1
ROUND	6	6	6	6	1	9	9	9	9	6
FLAT	6	6	6	6	1	9	1	1	6	1
OLD	6	6	6	6	1	9	1	1	6	1
YOUNG	6	6	6	6	1	1	1	1	6	1
GOOD	6	6	6	6	1	9	1	1	6	1
BAD	6	6	6	6	1	9	1	1	6	1
BIG	6	6	6	6	1	9	1	1	6	1
SHORT	6	6	6	6	1	9	1	1	6	1
HEAVY	6	6	6	6	1	9	1	1	6	1
SOFT	6	6	6	6	1	9	1	1	6	1
HAPPY	6	6	6	6	1	9	1	1	6	1
ANGRY	6	6	6	6	1	9	1	1	6	1
KNOW	6	6	1	6	6	1	1	6	1	6
WANT	6	6	1	6	6	1	1	6	1	6
LOVE	1	6	1	6	6	6	1	6	1	6
SEE/LOOK AT	6	1	1	6	6	1	1	6	1	6
SIT	1	6	1	6	6	9	9	9	9	9
WEAR	1	6	1	6	6	9	9	9	9	9
FIGHT	6	6	1	6	6	6	1	6	1	6
EAT	1	6	1	6	6	1	1	6	1	6
COME	1	1	1	6	6	6	1	6	1	6
COOK	1	6	6	1	6	1	1	6	1	6
BUILD	1	1	1	6	6	6	1	6	1	6
COUGH	1	6	6	1	6	1	1	6	1	6
HIT	1	6	1	6	6	6	1	6	1	6
WAVE	6	6	1	6	6	6	1	6	1	6
DIE	1	1	1	6	6	1	1	6	1	6
GIVE	1	1	1	6	6	1	1	6	1	6
BREAK	1	6	1	6	6	6	1	6	1	6

Appendix B. Parts of speech data coded for MDS analysis (cont'd)

	KA-M-SC-Aff	KA-M-BP-PL	KA-M-BP-Intens	KA-M-BP-0	KA-P-SC-O	KA-P-SC-Comp-Verb	KA-P-BP-TAM-Dir	KA-P-BP-TAM-Indir	KA-P-BP-0
WOMAN	1	6	6	1	1	6	6	6	1
BOY	1	6	6	1	1	6	6	6	1
FATHER	1	6	6	1	1	6	6	6	1
SISTER	1	6	6	1	1	6	6	6	1
DOG	1	6	6	1	1	6	6	6	1
BIRD	1	6	6	1	1	6	6	6	1
TREE	1	6	6	1	1	6	6	6	1
SEED	1	6	6	1	1	6	6	6	1
HAT	1	6	6	1	1	6	6	6	1
BED	1	6	6	1	1	6	6	6	1
ARM	1	6	6	1	1	6	6	6	1
FACE	1	6	6	1	1	6	6	6	1
RIVER	1	6	6	1	1	6	6	6	1
HOUSE	1	6	6	1	1	6	6	6	1
WATER	1	6	6	1	1	6	6	6	1
STONE	1	6	6	1	1	6	6	6	1
MALE	6	6	9	9	1	6	6	6	1
FEMALE	6	6	9	9	1	6	6	6	1
WHITE	6	6	1	6	1	6	6	6	1
RED	6	6	1	6	1	6	6	6	1
ROUND	1	6	9	9	6	6	6	6	1
FLAT	6	6	1	6	1	6	6	6	1
OLD	6	6	1	6	1	6	6	6	1
YOUNG	6	6	1	6	1	6	6	6	1
GOOD	6	6	1	6	1	6	6	6	1
BAD	6	6	1	6	1	6	6	6	1
BIG	6	1	1	6	1	6	6	6	1
SHORT	6	6	1	6	1	6	6	6	1
HEAVY	6	6	1	6	1	6	6	6	1
SOFT	6	6	1	6	1	6	6	6	1
HAPPY	6	6	1	6	1	6	6	6	1
ANGRY	6	6	1	6	1	6	6	6	1
KNOW	1	6	6	1	6	1	6	1	6
WANT	1	6	6	1	1	1	1	1	6
LOVE	1	6	6	1	1	6	1	6	6
SEE/LOOK AT	1	6	6	1	1	6	1	6	6
SIT	9	9	9	9	9	9	9	9	9
WEAR	9	9	9	9	9	9	9	9	9
FIGHT	1	6	6	1	1	1	1	1	6
EAT	1	6	6	1	1	6	1	6	6
COME	1	6	6	1	1	6	1	6	6
COOK	1	6	6	1	6	1	6	1	6
BUILD	1	6	6	1	1	1	1	1	6
COUGH	1	6	6	1	1	6	1	6	6
HIT	1	6	6	1	6	1	6	1	6
WAVE	1	6	6	1	1	6	1	6	6
DIE	1	6	6	1	1	1	1	1	6
GIVE	1	6	6	1	1	6	1	6	6
BREAK	1	6	6	1	6	1	6	1	6

Appendix C. MDS output table for points

	rank	correctYea	wrongYea	wrongNay	correctNay	volume	coord1D	coord2D
WOMAN	49	75	2	0	90	0.006000001	0.861697508	0.428549482
BOY	48	74	1	0	91	0.022000002	0.859375265	0.417361899
FATHER	46	72	0	0	94	0.021000002	0.834974071	0.405622173
SISTER	45	73	0	5	88	0.012000001	0.834398446	0.418269535
DOG	47	72	0	0	94	0.030000001	0.853451011	0.384299431
BIRD	44	70	0	0	96	0.032000002	0.817363695	0.377763505
TREE	39.5	69	0	0	97	0.047000002	0.797914134	0.33058484
SEED	39.5	69	0	0	97	0.048000004	0.797914134	0.33058484
HAT	39.5	69	0	0	97	0.047000002	0.797914134	0.33058484
BED	39.5	64	0	0	93	0.049000002	0.797914134	0.33058484
ARM	39.5	69	0	0	97	0.049000002	0.797914134	0.33058484
FACE	39.5	69	0	0	97	0.050000001	0.797914134	0.33058484
RIVER	39.5	64	0	0	90	0.050000001	0.797914134	0.33058484
HOUSE	39.5	69	0	0	97	0.050000001	0.797914134	0.33058484
WATER	35	68	1	0	97	0.025	0.792363729	0.321871172
STONE	34	69	0	0	97	0.011000001	0.769228641	0.286396204
MALE	31.5	42	2	0	67	0.002	0.33050118	-0.467034192
FEMALE	31.5	44	2	0	67	0.015000001	0.33050118	-0.467034192
WHITE	30	71	0	2	91	0.100000001	0.171691973	-0.508718317
RED	33	76	1	3	85	0.018000001	0.339187447	-0.515571356
ROUND	18	50	1	1	75	0.003	-0.321879643	-0.817196723
FLAT	27	62	0	0	86	0.027000001	0.10800206	-0.853864506
OLD	24.5	70	1	1	92	0.150000006	0.001233645	-0.720621782
YOUNG	28	68	1	1	95	0.026000001	0.108121724	-0.854013585
GOOD	24.5	69	2	2	91	0.149000004	0.001233645	-0.720621782
BAD	20	65	5	5	89	0.213000014	-0.101191287	-0.71706093
BIG	29	71	1	0	92	0.032000002	0.109490557	-0.868232146
SHORT	26	70	0	0	93	0.181000009	0.017681246	-0.747266058
HEAVY	21	72	1	2	85	0.045000002	-0.098314302	-0.55680758
SOFT	22	73	1	0	91	0.014	-0.08870984	-0.542168107
HAPPY	19	62	0	3	90	0.003	-0.284728932	-0.532321416
ANGRY	23	67	2	3	88	0.002	-0.08686224	-0.511127201
KNOW	14	63	0	2	104	0.005	-0.763423849	0.212456871
WANT	16	64	0	3	102	0.018000001	-0.723274315	0.223196688
LOVE	17	62	2	4	98	0.003	-0.720340106	0.226282023
SEE/LOOK AT	15	65	0	1	103	0.020000001	-0.726102476	0.224426089
SIT	4	59	1	0	95	0.006000001	-0.859664189	0.190334549
WEAR	3	54	0	0	89	0.007	-0.859962073	0.189018745
FIGHT	8	65	2	1	101	0.042000003	-0.808105513	0.379253998
EAT	9	64	1	2	99	0.011000001	-0.789116407	0.287659753
COME	1	67	7	1	94	0.002	-0.913218145	0.27430276
COOK	13	65	5	1	95	0.004	-0.773803421	0.490559158
BUILD	7	64	0	0	93	0.017000001	-0.823639237	0.371566355
COUGH	10	63	3	3	100	0.012000001	-0.789116407	0.287659753
HIT	2	67	4	0	98	0.012000001	-0.899269646	0.282302936
WAVE	12	55	2	0	85	0.014	-0.777703801	0.437822889
DIE	5	69	2	1	97	0.003	-0.837063773	0.392426895
GIVE	11	66	0	3	100	0.010000001	-0.785703816	0.289630448
BREAK	6	65	1	0	103	0.037	-0.829541173	0.31672256

Appendix D. MDS output table for cutting lines

	correctYea	wrongYea	wrongNay	correctNay	PRE	normVector1D	normVector2D	midpoints
KH.R.SC.0	17	0	0	30	1	0.805007101	0.593265174	-0.103897367
KH.R.SC.VarStr	31	0	1	15	0.933333333	0.135342226	0.990798911	0.384659862
KH.R.BP.PGN	17	0	0	30	1	0.805007101	0.593265174	-0.103897367
KH.R.BP.0	31	0	1	15	0.933333333	0.135342226	0.990798911	0.384659862
KH.M.SC.0	31	0	1	15	0.933333333	0.135342226	0.990798911	0.384659862
KH.M.SC.VarStr	17	0	0	30	1	0.805007101	0.593265174	-0.103897367
KH.M.BP.SPRL1	1	0	0	46	1	0.083924075	-0.996472152	0.865474831
KH.M.BP.SPRL2	31	1	0	15	0.9375	0.886533834	-0.462663767	0.548464427
KH.M.BP.0	16	0	0	31	1	0.759709206	0.650262964	-0.052589374
KH.P.SC.0	31	0	1	15	0.933333333	0.135342226	0.990798911	0.384659862
KH.P.SC.Cop	17	0	0	30	1	0.805007101	0.593265174	-0.103897367
KH.P.BP.TAM1	17	0	0	30	1	0.815400825	-0.578896791	-0.322512825
KH.P.BP.TAM2	14	0	1	32	0.933333333	0.61189348	-0.790940181	0.245815811
KH.P.BP.0	16	0	0	31	1	0.759709206	0.650262964	-0.052589374
KS.R.SC.0	15	0	3	27	0.833333333	0.905904266	-0.42348254	0.554941479
KS.R.SC.Aff	28	3	0	15	0.833333333	0.901715792	-0.432329307	0.553038699
KS.R.BP.Class	15	0	0	27	1	0.759709206	0.650262964	-0.052589374
KS.R.BP.Num	14	1	0	27	0.928571429	0.131411632	0.991327889	0.384872399
KS.R.BP.0	27	0	0	15	1	0.759709206	0.650262964	-0.052589374
KS.M.SC.0	12	0	1	17	0.923076923	0.833042122	0.553209566	-0.403280106
KS.M.SC.Suff	21	1	0	3	0.75	0.069839772	0.997558222	-0.839075576
KS.P.SC.0	23	0	2	22	0.909090909	0.858824074	0.512270642	-0.362859967
KS.P.SC.Cop	28	1	1	15	0.875	0.798757173	-0.601653537	0.237371388
KS.P.BP.TAMext	23	0	2	22	0.909090909	0.858824074	0.512270642	-0.362859967
KS.P.BP.0	28	1	1	15	0.875	0.798757173	-0.601653537	0.237371388
JI.R.SC.0	16	0	0	15	1	0.815400825	-0.578896791	0.144504531
JI.R.SC.Aff	15	0	0	16	1	0.759709206	0.650262964	-0.052589374
JI.R.BP.NGC	18	0	0	29	1	0.910031919	0.414538184	-0.151896832
JI.R.BP.0	29	0	0	18	1	0.848618008	0.529006122	-0.236404902
JI.M.SC.0	14	0	0	15	1	0.815400825	-0.578896791	-0.322512825
JI.M.SC.Aff	31	0	0	14	1	0.077290281	-0.997008632	0.132589649
JI.M.BP.NGCagr	30	0	0	15	1	0.815400825	-0.578896791	-0.322512825
JI.M.BP.0	15	0	0	30	1	0.815400825	-0.578896791	-0.322512825
JI.P.SC.0	15	1	0	29	0.933333333	0.392207382	0.919876823	-0.10554094
JI.P.SC.Aff	16	0	0	29	1	0.759709206	0.650262964	0.359377534
JI.P.SC.LV	14	1	0	30	0.928571429	0.140628632	-0.990062416	-0.307058682
JI.P.BP.TAM.abl	1	0	0	44	1	0.903766448	-0.428025943	-0.938371886
JI.P.BP.TAM.peri	14	1	0	30	0.928571429	0.140628632	-0.990062416	-0.307058682
JI.P.BP.0	30	0	0	15	1	0.815400825	-0.578896791	-0.322512825
MA.R.BP.NPArt	17	0	0	31	1	0.966883805	-0.255216983	0.452026892
MA.R.BP.SpecArt	29	0	0	19	1	0.871261763	-0.490818643	0.181535394
MA.R.BP.NClass	29	0	0	19	1	0.871261763	-0.490818643	0.181535394
MA.R.BP.0	19	0	0	29	1	0.871261763	-0.490818643	0.181535394
MA.M.SC.REL	31	0	0	17	1	0.966883805	-0.255216983	0.452026892
MA.M.SC.REL.Prep	17	0	0	31	1	0.966883805	-0.255216983	0.452026892
MA.M.SC.PossPron	17	0	0	31	1	0.966883805	-0.255216983	0.452026892
MA.M.BP.Intens	12	1	0	35	0.916666667	0.105814136	-0.994385925	0.499294263

Appendix D. MDS output table for cutting lines (cont'd)

	correctYea	wrongYea	wrongNay	correctNay	PRE	normVector1D	normVector2D	midpoints
MA.M.BP.0	35	0	1	12	0.916666667	0.105869247	-0.994380059	0.499292599
MA.P.SC.0	19	0	0	29	1	0.871261763	-0.490818643	0.181535394
MA.P.SC.Cop	29	0	0	19	1	0.871261763	-0.490818643	0.181535394
MA.P.BP.ConcPron	31	0	0	17	1	0.966883805	-0.255216983	0.452026892
MA.P.BP.T1	3	1	1	43	0.5	0.996973736	-0.077739114	-0.867734984
MA.P.BP.T2	11	4	0	33	0.636363636	0.148875981	-0.988855875	-0.319300761
MA.P.BP.T3	1	1	1	45	0	0.906657705	-0.421867048	-0.929610446
MA.P.BP.T4	31	0	0	17	1	0.815400825	-0.578896791	-0.322512825
TO.R.SC.0	15	0	0	30	1	0.759709206	0.650262964	-0.052589374
TO.R.SC.Aff	30	0	0	15	1	0.759709206	0.650262964	-0.052589374
TO.R.BP.Num	15	0	0	30	1	0.759709206	0.650262964	-0.052589374
TO.R.BP.Pron	4	0	0	41	1	0.551009719	0.834498825	0.79789426
TO.R.BP.PLmark	5	0	0	40	1	0.72049683	0.69345823	0.868118196
TO.R.BP.SubjIndN	5	0	0	40	1	0.72049683	0.69345823	0.868118196
TO.R.BP.0	30	0	0	15	1	0.759709206	0.650262964	-0.052589374
TO.M.SC.0	31	0	0	16	1	0.815400825	-0.578896791	-0.322512825
TO.M.SC.Aff	17	0	0	30	1	0.832039125	0.55471695	0.003333235
TO.M.SC.RelMark	30	0	0	17	1	0.832039125	0.55471695	0.003333235
TO.P.BP.TAM	30	0	0	17	1	0.832039125	0.55471695	0.003333235
TO.P.BP.0	17	0	0	30	1	0.832039125	0.55471695	0.003333235
QE.R.SC.0	16	0	0	18	1	0.906241601	-0.422760169	0.011581476
QE.R.SC.Aff	21	0	0	16	1	0.759709206	0.650262964	-0.052589374
QE.M.SC.0	31	0	0	18	1	0.845659021	-0.533723543	0.104845334
QE.M.SC.Suff	37	0	0	12	1	0.439234795	-0.898372303	0.499757669
QE.M.BP.SPRL	16	0	0	33	1	0.657924726	-0.753083697	0.286430384
QE.M.BP.0	36	0	0	13	1	0.710980511	-0.703211712	0.342388845
QE.P.SC.0	21	0	0	28	1	0.746366578	-0.665535071	0.302672077
QE.P.SC.Cop	31	0	0	18	1	0.845659021	-0.533723543	0.104845334
QE.P.BP.TAM	21	0	0	28	1	0.746366578	-0.665535071	0.302672077
QE.P.BP.TAMperi	31	0	0	18	1	0.845659021	-0.533723543	0.104845334
CR.R.SC.0	16	0	0	31	1	0.759709206	0.650262964	-0.052589374
CR.R.SC.Aff	31	0	0	16	1	0.759709206	0.650262964	-0.052589374
CR.R.BP.Num	2	0	0	45	1	0.758010496	0.652242354	0.91607827
CR.R.BP.Dim	16	0	0	31	1	0.759709206	0.650262964	-0.052589374
CR.R.BP.0	31	0	0	16	1	0.759709206	0.650262964	-0.052589374
CR.M.SC.Attach	14	0	0	33	1	0.077290281	-0.997008632	0.132589649
CR.M.SC.Aff	36	0	0	11	1	0.778488432	-0.627658953	-0.790927946
CR.M.SC.AffLgr	11	0	0	36	1	0.778488432	-0.627658953	-0.790927946
CR.M.BP.CMPR	20	0	0	27	1	0.218296833	0.975882418	0.064665113
CR.M.BP.0	27	0	0	20	1	0.218296833	0.975882418	0.064665113
CR.P.SC.0	17	0	0	30	1	0.815400825	-0.578896791	-0.322512825
CR.P.SC.DURAff	14	0	0	33	1	0.077290281	-0.997008632	0.132589649
CR.P.SC.Cop	30	0	0	17	1	0.815400825	-0.578896791	-0.322512825
CR.P.BP.TAM	31	0	0	16	1	0.759709206	0.650262964	-0.052589374
CR.P.BP.TAMperi	30	0	0	17	1	0.815400825	-0.578896791	-0.322512825
KI.R.SC.0	18	0	0	26	1	0.910031919	0.414538184	-0.151896832
KI.R.SC.Aff	26	0	0	18	1	0.910031919	0.414538184	-0.151896832

Appendix D. MDS output table for cutting lines (cont'd)

	correctYea	wrongYea	wrongNay	correctNay	PRE	normVector1D	normVector2D	midpoints
KI.R.BP.Num.Suff	2	0	0	42	1	0.758010496	0.652242354	0.91607827
KI.R.BP.Num.Peri	14	2	0	28	0.857142857	0.134894412	0.990859979	0.384684387
KI.R.BP.Num.Vind	6	0	0	38	1	0.726753379	0.686898483	0.851460168
KI.R.BP.0	28	0	0	16	1	0.759709206	0.650262964	-0.052589374
KI.M.SC.0	15	0	0	33	1	0.108156838	-0.994133843	0.498570448
KI.M.SC.GEN	16	0	0	32	1	0.759709206	0.650262964	0.359377534
KI.M.SC.REL	17	0	0	31	1	0.845659021	-0.533723543	0.104845334
KI.M.BP.CMPR	14	0	0	34	1	0.03816815	-0.999271331	0.493397837
KI.M.BP.0	35	0	0	14	1	0.03816815	-0.999271331	0.493397837
KI.P.SC.0	32	1	0	15	0.9375	0.905462452	-0.424426376	0.57581991
KI.P.SC.PlusNoun	0	0	1	47	0	0.995343772	0.096388671	0.903267054
KI.P.SC.FOC	15	0	0	33	1	0.750974118	0.66033164	0.778518423
KI.P.BP.TAMIndArg	17	0	0	31	1	0.845659021	-0.533723543	0.104845334
KI.P.BP.0	31	0	0	17	1	0.845659021	-0.533723543	0.104845334
EN.R.SC.0	19	2	0	28	0.894736842	0.836046157	0.548659115	-0.360463655
EN.R.SC.Aff	30	3	0	16	0.842105263	0.894682439	-0.446702734	0.549675992
EN.R.SC.Peri	0	0	2	47	0	0.997660913	0.068357168	0.893144248
EN.R.BP.NumAbl	1	0	0	48	1	0.080235799	0.996775911	0.493979893
EN.R.BP.NumSuff	15	1	0	33	0.933333333	0.13658569	0.99062826	0.38459134
EN.R.BP.0	33	0	0	16	1	0.759709206	0.650262964	0.359377534
EN.M.SC.0	28	0	1	20	0.95	0.705045934	-0.70916164	0.301908103
EN.M.SC.Aff	36	2	0	11	0.846153846	0.004674728	-0.999989073	0.530793426
EN.M.SC.REL	17	0	0	32	1	0.815400825	-0.578896791	-0.322512825
EN.M.BP.CMPRAbl	0	0	2	47	0	0.994291391	0.106698784	0.906817239
EN.M.BP.CMPRSuff	10	3	0	36	0.7	0.101960971	-0.9947884	0.500063737
EN.M.BP.CPRSuff2	0	0	1	48	0	0.148835634	0.988861949	-0.858103443
EN.M.BP.CMPRPeri	1	0	0	48	1	0.122816239	0.992429429	-0.849346925
EN.M.BP.SPRLAbl	0	0	2	47	0	0.994291391	0.106698784	0.906817239
EN.M.BP.SPRLSuff	11	2	0	36	0.818181818	0.105237598	-0.994447107	0.499311583
EN.M.BP.SPRLSuff2	0	0	1	48	0	0.148835634	0.988861949	-0.858103443
EN.M.BP.0	35	0	0	14	1	0.03816815	-0.999271331	0.493397837
EN.P.SC.0	17	0	0	32	1	0.815400825	-0.578896791	-0.322512825
EN.P.SC.Cop	18	0	0	31	1	0.157747437	-0.987479491	-0.19994926
EN.P.SC.CopArt	14	0	0	35	1	0.746614471	0.665256966	0.812687789
EN.P.BP.TAbl	11	4	0	34	0.636363636	0.35504321	-0.934849891	-0.467402502
EN.P.BP.Taff	3	0	3	43	0.5	0.759653708	-0.650327797	-0.872386957
EN.P.BP.TPeri	32	0	0	17	1	0.815400825	-0.578896791	-0.322512825
EN.P.BP.P+N.Suff	17	0	0	32	1	0.815400825	-0.578896791	-0.322512825
EN.P.BP.P+N.Peri	32	0	0	17	1	0.815400825	-0.578896791	-0.322512825
IN.R.SC.0	16	0	2	31	0.888888889	0.888541909	-0.458795461	0.546715708
IN.R.SC.Aff	33	0	0	16	1	0.759709206	0.650262964	0.359377534
IN.R.BP.CaseNum	32	0	0	17	1	0.815400825	-0.578896791	-0.322512825
IN.R.BP.0	17	0	0	32	1	0.815400825	-0.578896791	-0.322512825
IN.M.SC.0	16	0	0	33	1	0.077290281	-0.997008632	0.132589649
IN.M.SC.Aff	34	0	1	14	0.928571429	0.679071109	-0.734072496	0.314391842
IN.M.BP.CMPR	14	0	0	35	1	0.03816815	-0.999271331	0.493397837
IN.M.BP.GendAgr	1	0	1	47	0.5	0.112080302	0.993699153	-0.84931586

Appendix D. MDS output table for cutting lines (cont'd)

	correctYea	wrongYea	wrongNay	correctNay	PRE	normVector1D	normVector2D	midpoints
IN.M.BP.NumAgr	1	0	0	48	1	0.083924075	-0.996472152	0.865474831
IN.M.BP.CaseAgr	33	0	0	16	1	0.759709206	0.650262964	0.359377534
IN.M.BP.0	16	0	0	33	1	0.759709206	0.650262964	0.359377534
IN.P.SC.0	15	2	0	32	0.866666667	0.146680095	-0.989183982	-0.306770662
IN.P.SC.Comp	1	0	1	47	0.5	0.840059075	-0.542494931	-0.916119457
IN.P.SC.Cop	32	0	0	17	1	0.815400825	-0.578896791	-0.322512825
IN.P.BP.Iter	1	1	1	46	0	0.953009227	0.302941269	-0.76640673
IN.P.BP.Aux	3	1	1	44	0.5	0.941653131	-0.33658488	-0.896301325
IN.P.BP.IndNum	2	0	0	47	1	0.840928597	-0.541146093	-0.916224298
IN.P.BP.IndGend	12	2	0	35	0.833333333	0.26358795	-0.964635368	-0.408055624
IN.P.BP.Suppletive	3	1	2	43	0.4	0.950644026	-0.310283639	-0.892136433
IN.P.BP.FullAkt	15	2	0	32	0.866666667	0.146680095	-0.989183982	-0.306770662
IN.P.BP.PartAkt	1	0	1	47	0.5	0.840059075	-0.542494931	-0.916119457
IN.P.BP.RestrAkt	32	0	0	17	1	0.815400825	-0.578896791	-0.322512825
KA.R.SC.0	24	1	2	6	0.571428571	0.145426729	-0.989369024	-0.403838328
KA.R.SC.Aff	30	0	0	16	1	0.759709206	0.650262964	0.359377534
KA.R.BP.PL	31	0	0	15	1	0.815400825	-0.578896791	-0.322512825
KA.R.BP.0	15	0	0	31	1	0.815400825	-0.578896791	-0.322512825
KA.M.SC.0	30	0	1	16	0.9375	0.807219254	-0.590251706	0.228067302
KA.M.SC.Aff	31	0	1	15	0.933333333	0.609497902	-0.792787682	0.2454613
KA.M.BP.PL	1	0	0	46	1	0.083924075	-0.996472152	0.865474831
KA.M.BP.Intens	13	0	0	31	1	0.077290281	-0.997008632	0.132589649
KA.M.BP.0	31	0	0	13	1	0.077290281	-0.997008632	0.132589649
KA.P.SC.O	41	1	1	4	0.6	0.919650781	0.392737115	-0.61577581
KA.P.SC.CompVerb	6	1	2	38	0.625	0.940998987	-0.338409378	-0.882509908
KA.P.BP.TAMDir	11	3	0	33	0.727272727	0.137246678	-0.9905369	-0.318698564
KA.P.BP.TAMIndir	6	1	2	38	0.625	0.940998987	-0.338409378	-0.882509908
KA.P.BP.0	32	0	0	15	1	0.815400825	-0.578896791	-0.322512825

REFERENCES

Anderson, Lloyd B. 1974. Distinct sources of fuzzy data: ways of integrating relatively discrete and gradient aspects of language, and explaining grammar on the basis of semantic fields. In Roger W. Shuy & Charles-James N. Bailey (eds.), *Towards tomorrow's linguistics*, 50–64. Washington, D.C.: Georgetown University Press.

Anderson, Lloyd B. 1982. The 'perfect' as a universal and as a language-particular category. In Paul Hopper (ed.), *Tense-aspect: Between semantics and pragmatics*, 227–64. Amsterdam: John Benjamins.

Anderson, Lloyd B. 1986. Evidentials, paths of change, and mental maps: Typologically regular asymmetries. In Wallace Chafe & Johanna Nichols (eds.), *Evidentiality: The linguistic encoding of epistemology*, 273–312. Norwood: Ablex.

Anderson, Lloyd B. 1987. Adjectival morphology and semantic space. In Barbara Need, Eric Schiller & Ann Bosch (eds.), *Papers from the 23rd Annual Regional Meeting of the Chicago Linguistic Society, Part One: The General Session*, 1–17. Chicago: Chicago Linguistic Society.

Auwera, Johan van der & Vladimir A. Plungian. 1998. Modality's semantic map. *Linguistic Typology* 2. 79–124.

Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Bickel, Balthasar. 2010. Capturing particulars and universals in clause linkage: A multivariate analysis. In Isabelle Bril (ed.), *Clause-hierarchy and clause-linking: The syntax and pragmatics interface*, 51-101. Amsterdam: John Benjamins.

Borg, Ingwer & Patrick Groenen. 1997. *Modern multidimensional scaling: Theory and applications*. New York: Springer.

Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form.* Amsterdam: Benjamins.

Bybee, Joan L., Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.

Chafe, Wallace L. 1970. *Meaning and the structure of language*. Chicago: University of Chicago Press.

Chafe, Wallace L. (ed.). 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, New Jersey: Ablex.

Childs, George Tucker. 1995. *A grammar of Kisi: a Southern Atlantic language*. Berlin: de Gruyter.

Childs, George Tucker. 2000. *A dictionary of Kisi*. Cologne: Rudiger Koppe.

Comrie, Bernard. 1975. Polite plurals and predicate agreement. *Language* 51(2). 406-18.

Comrie, Bernard. 1976. *Aspect*. Cambridge: Cambridge University Press.

Comrie, Bernard. 1989. *Language universals and linguistic typology*. 2nd edn. Chicago: University of Chicago Press.

Corbett, Greville G. 2000. *Number*. Cambridge: Cambridge University Press.

Croft, William. 1991. *Syntactic categories and grammatical relations: The cognitive organization of information*. Chicago: University of Chicago Press.

Croft, William. 1996. 'Markedness' and 'universals': from the Prague school to typology. In Kurt R. Jankowsky (ed.), *Multiple perspectives on the historical dimensions of language*, 15-21. Münster: Nodus.

Croft, William. 2001. *Radical construction grammar.* Oxford: Oxford University Press.

Croft, William. 2003. *Typology and universals*. 2nd edn. Cambridge: Cambridge University Press.

Croft, William. 2012. *Verbs: Aspect and causal structure*. Oxford: Oxford University Press.

Croft, William (to appear). *Morphosyntax*. Cambridge: Cambridge University Press.

Croft, William, Hava Bat-Zeev Shyldkrot & Suzanne Kemmer. 1987. Diachronic semantic processes in the middle voice. In Anna Giacolone Ramat, Onofrio Carruba & Guiliano Bernini (eds.), *Papers from the 7th International Conference on Historical Linguistics*, 179–192. Amsterdam: John Benjamins.

Croft, William & Keith T. Poole. 2008. Inferring universals from grammatical variation: Multidimensional scaling for typological analysis. *Theoretical Linguistics* 34(1). 1-37.

Cyffer, Norbet & John Hutchison. 1990. *Dictionary of the Kanuri language*. Providence: Foris.

Cysouw, Michael. 2007. New approaches to cluster analysis of typological indices. In Reinhard Köhler & Peter Grzbek (eds.), *Exact methods in the study of language and text*, 61-76. Berlin: Mouton de Gruyter.

Dahl, Östen. 1985. *Tense and aspect systems*. Oxford: Basil Blackwell.

Dixon, Robert M. W. 1979. Ergativity. *Language* 55(1). 59-138.

Dixon, Robert M. W. 1982. *Where have all the adjectives gone?* Berlin: Mouton de Gruyter.

Du Bois, John W. 1985. Competing motivations. In John Haiman (ed.), *Iconicity in syntax*, 343-65. Amsterdam: John Benjamins.

Fillmore, Charles J. 1970. The grammar of *hitting* and *breaking*. In Roderick A. Jacobs & Peter S. Rosenbaum (eds.), *Readings in English Transformational Grammar*, 120–133. Waltham, MA: Ginn.

Givón, Talmy. 1979. *On understanding grammar*. New York: Academic Press.

Greenberg, Joseph H. 1960. The general classification of Central and South American languages. In Anthony Wallace (ed.), *Men and Cultures: Selected papers of the 5th International Congress of Anthropological and Ethnological Sciences*, 791-94. Philadelphia, University of Pennsylvania Press.

Greenberg, Joseph H. 1966/2005. *Language universals: With special reference to feature hierarchies*. New York: Mouton de Gruyter.

Greenberg, Joseph H. 1987. *Language in the Americas*. Stanford: Stanford University Press.

Greenberg, Joseph H. 1990. Some universals of grammar with particular reference to the order of meaningful elements. In Keith Denning & Suzanne Kemmer (eds.), *On language: Selected writings of Joseph H. Greenberg*, 40-70. Stanford: Stanford University Press.

Haspelmath, Martin. 1997a. *Indefinite pronouns*. Oxford: Oxford University Press.

Haspelmath, Martin. 1997b. *From space to time: Temporal adverbials in the world's languages*. München: Lincom Europa.

Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In Michael Tomasello (ed.), *The new psychology of language*, vol. 2, 211–42. Mahwah, NJ: Lawrence Erlbaum Associates.

Hengeveld, Kees. 1992. *Non-verbal predication: Theory, typology, diachrony.* Berlin: Mouton de Gruyter.

Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251-299.

Hopper, Paul J. & Sandra A. Thompson. 1984. The discourse basis for lexical categories in universal grammar. *Language* 60(4). 703-752.

Hutchison, John. 1981. *The Kanuri language: A reference grammar*. Madison: University of Wisconsin, Madison.

Jakobson, Roman. 1932/1984. Structure of the Russian verb. In Linda R. Waugh & Morris Hale (eds.), *Russian and Slavic grammar: Studies, 1931-1981* (Janua Linguarum, Series Maior 106), 1-14. The Hague: Mouton.

Jakobson, Roman. 1939/1984. Zero sign. In Linda R. Waugh & Morris Hale (eds.), *Russian and Slavic grammar: Studies, 1931-1981* (Janua Linguarum, Series Maior 106), 151-160. The Hague: Mouton.

Kemmer, Suzanne. 1983. *The middle voice*. (Typological Studies in Language, 23.) Amsterdam: John Benjamins.

Kilian-Hatz, Christa. 2003. *Khwe dictionary*. Cologne: Rüdiger Koppe.

Kilian-Hatz, Christa. 2008. *A grammar of modern Khwe (Central Khoisan)*. Cologne: Rüdiger Koppe.

Lehmann, Christian & Edith Moravcsik. 2000. Nouns. In Geert Booij, Christian Lehmann & Joachim Mugdan (eds.), *Morphology: An international handbook on inflection and word-formation*, vol. 1, 732-57. Berlin: de Gruyter.

Langacker, Ronald W. 1987. Nouns and verbs. *Language* 63(1). 53-94.

Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.

Levin, Beth (in press). Verb classes within and across languages. In Bernard Comrie and Andrej Malchukov (eds.), *Valency classes: A comparative handbook*. Berlin: De Gruyter.

Levinson, Stephen C. & Sérgio Meira. 2003. 'Natural concepts' in the spatial topological domain—adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language* 79(3). 485-516.

Lichtenberk, Frantisek. 2008a. *A dictionary of Toqabaqita*. Canberra: Pacific Linguistics.

Lichtenberk, Frantisek. 2008b. *A grammar of Toqabaqita*. Berlin: de Gruyter.

Martin, Jack B. 2011. *A grammar of Creek (Muskogee)*. Lincoln: University of Nebraska Press.

Martin, Jack B. & Margaret M. Mauldin. 2000. *A dictionary of Creek/Muskogee*. Lincoln: University of Nebraska Press.

Nichols, Johanna. 2004. *Ingush-English and English-Ingush dictionary*. New York: RutledgeCurzon.

Nichols, Johanna. 2011. *Ingush grammar*. Berkeley: University of California Press.

Pensalfini, Robert. 1997. *Jingulu grammar, dictionary, and texts*. Cambridge: Massachusetts Institute of Technology doctoral dissertation.

Pensalfini, Robert. 2003. *A grammar of Jingulu*. Canberra: Pacific Linguistics.

Poole, Keith T. 2000. Non-parametric unfolding of binary choice data. *Political Analysis* 8. 211–237.

Poole, Keith T. 2005. *Spatial models of parliamentary voting*. Cambridge: Cambridge University Press.

Rosch, Eleanor. 1978. Principles of categorization. In Eleanor Rosch & Barbara B. Lloyd (eds), *Cognition and categorization*, 27–48. Hillsdale, NJ: Lawrence Erlbaum.

Ross, John R. 1972. The category squish: Endstation Hauptwort. In Paul M. Peranteau, Judith N. Levi & Gloria C. Phares (eds.), *Proceedings of the Eighth Regional Meeting of the Chicago Linguistic Society*, 316-28. Chicago: Chicago Linguistic Society.

Ruhlen, Merritt. 1991. The Amerind phylum and the prehistory of the New World. In Sydney M. Lamb & E. Douglas Mitchell (eds.), *Sprung from some common source*, 328-50. Stanford: Stanford University Press.

Schachter, Paul & Timothy Shopen. 2007. Parts of speech systems. In Timothy Shopen (ed.), *Language typology and syntactic description,* vol. 1: *Clause structure*, 2nd edn., 1-60. Cambridge: Cambridge University Press.

Searle, John R. 1969. *Speech acts: an essay in the philosophy of language*. Cambridge: Cambridge University Press.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Robert M. W. Dixon (ed.), *Grammatical categories in Australian languages*, 112-171. New Jersey: Humanities Press.

Smith, Carlota S. 1997. *The parameter of aspect.* Dordrecht: Kluwer Academic Publishers.

Stanley, Thomas. 1687. *The history of philosophy*. Garland Pub.

Stassen, Leon. 1997. *Intransitive predication*. Oxford: Clarendon Press.

Stebbins, Tonya N. 2011. *Mali (Baining) grammar*. Canberra: Pacific Linguistics.

Stebbins, Tonya N. 2012. *Mali (Baining) dictionary*. Canberra: Asia-Pacific Linguistics.

Trubetzkoy, Nikolai. 1931. Die phonologischen Systeme. *Travaux du Cercle Linguistique de Prague* 4. 96-116.

Trubetzkoy, Nikolai. 1939/1969. *Principles of phonology*. (Translation of *Grundzüge der Phonologie*, Prague, 1939). Berkeley and Los Angeles: University of California Press.

Vendler, Zeno. 1967. *Linguistics in philosophy*. Ithaca: Cornell University Press.

Weber, David J. 1989. *A grammar of Huallaga (Huánuco) Quechua*. Berkeley: University of California.

Weber, David J., Félix C. Zambrano, Teodoro C. Villar & Marlene B. Dávila. 2008. *Rimaycuna: Quechua de Huánuco* (Serie Lingüística Peruana 48). 2nd edn. Lima: Instituto Lingüístico de Verano.

Wetzer, Harrie. 1996. *The typology of adjectival predication*. Berlin: Mouton de Gruyter.