5-1-2016

# Unexpected Accompaniment: Cognitive Mechanisms for Language-Music Mismatch in Time

Hui-Chieh Hsu

**Hui-Chieh Hsu**
*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

**Sherman Wilcox**

**David Bashwiner**

**UNEXPECTED ACCOMPANIMENT:**
**COGNITIVE MECHANISMS FOR LANGUAGE-MUSIC MISMATCH IN TIME**

*by*

**HUI-CHIEH HSU**
**B.B.A., International Business, National Taiwan University, 2013**

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**

My gratitude goes to Drs. William Croft, Sherman Wilcox, and David Bashwiner for their insightful inputs and patient guidance, to Dr. I-wen Su for the academic rigor I will always benefit from, to the audiences at ICLC-13 and MIDI for their valuable comments, to my colleagues for the stimulating discussions, to my family and friends for their care and support, to Richard Wagner for the ethereal beauty he left for humanity, and to Albuquerque, she who made me understand what truly matters.

**UNEXPECTED ACCOMPANIMENT:**
**COGNITIVE MECHANISMS FOR LANGUAGE-MUSIC MISMATCH IN TIME**

by

**Hui-Chieh Hsu**
**B.B.A., International Business, National Taiwan University, 2013**
**M.A., Linguistics, University of New Mexico, 2016**

## Abstract

Extensive research has been carried out for the role played by anticipation in human cognition (Brône 2012; Coulson 2005; Huron 2006). Employing the terms "anticipation" and "expectation" interchangeably, these studies emphasize only the temporal sequentiality between elements taking place at different points in time, a relation parallel to the progression of time. In modes of communication where simultaneous events are possible, however, there exists a similar but distinct relation between elements simultaneously represented in time. This is expectation, a relation perpendicular to the progression of time. Ubiquitous and indispensable for our understanding of the most ordinary, the relation of expectation has hardly been recognized let alone analyzed in the literature.

In view of this, in the present study I contemplate on the nature of expectation, identifying and integrating in a coherent way theoretical

notions relevant to the analysis of expectation, including among others the progression of time, conceptual distance (Gärdenfors 2000), processes of conceptual integration and disintegration (Fauconnier and Turner 2002), construction (Croft 2001; Langacker 2008), and intention. Analyzed for illustration are two passages from Wagner's music drama *Siegfried*: While in the first passage Mime sings words of care and love to music of complaint, in the second, a honeyed tune is heard accompanying the murderous intentions in Mime's words. The cross-modal mismatch in both cases stretches the conceptual distance between Mime's words and music, imposing tension on the listener's expectation that simultaneously represented language and music should "support" each other. Meanwhile, elements from the language and the music inputs, as well as the unexpected mismatch, are projected into a blend, where new meanings emerge and the mismatch is made sense of. Drawing upon Clark's (1996) insight, the multiplex layers of pretense and coatings of meaning in the convoluted social-interactional dynamics in the two passages are also disentangled and laid out.

The analysis of the two passages with cross-modal mismatch underlines the centrality of the hitherto overlooked relation of expectation, and that, much like unfulfilled anticipation, unfulfilled

expectation also yields strong emotional effects. The stark contrast between the ease of our understanding of the two passages and the complexity of the analysis further highlights the agility of the cognitive processing that allows us to make sense of the world we are exposed to.

# Table of Contents

# List of Figures

# List of Tables

*Expect: Latin* exspectāre, ex- *"out"* + spectare *"look," to look out for.*

*Anticipate: Latin* anticipare, ante- *"before"* + capers *"take," to act in advance*.

# 1. MIME'S COMPLAINT.

The First Act of *Siegfried*, the second day of Wagner's epic music drama *der Ring des Nibelungen*, starts with a conversation between Siegfried, the namesake protagonist, and Mime, a Nibelung dwarf who is after the Ring and who has raised Siegfried only for that purpose. In the First Scene of the Act, the listener hears "Als zullendes Kind zog ich dich auf," a passage where Mime narrates in words how much love he has for Siegfried and how much effort he has devoted and is still willing to devote to taking care of him. Curiously enough, the leitmotif representing "Mime's complaint" (Donington 1974) prevails in the music Mime sings his words to. In other words, there is a mismatch between the semantics of Mime's words and the accompanying tune. Given the context of the music drama and basic human cognitive ability, the listener is usually capable of interpreting the entire passage as conveying Mime's dishonesty: Mime is only saying nice things to Siegfried for the sake of the Ring, an evil intention his music unmasks. Indeed, as will be revealed at a later stage of the music drama, Mime actually attempts to kill Siegfried after his goal has been fulfilled.

Intuitive as it seems, the way the listener is able to make sense of Mime's real intention out of this mismatch is really a result of some marvelous cognitive operations. For instance, the listener not only has to understand first and foremost what the language and the music mean, respectively, but also needs to further compare

the linguistic meaning with the musical meaning in order for the two to clash and for new meaning to emerge. Among these, most fundamentally, the listener has to make a comparison: They need to know what ought to be compared to which, when a comparison should take place, and so on. This, however, happens to be what has long been overlooked. In the present study, I argue that this key comparison hinges on expectation. In the following sections, I elaborate on the idea of expectation, especially against the background of studies on other seemingly similar cognitive mechanisms, before demonstrating how expectation can be captured and analyzed. I then show how the above *Siegfried* example, as well as another similar but distinct passage, can be analyzed in this light.

## 2. ANTICIPATION AND EXPECTATION.

Extensive research has been carried out on the role anticipation plays in human cognition, including in linguistics, cognitive science, psychology, musicology, etc. To begin with, in their study of non-literal uses of language, such as irony and humor, cognitive linguists have alluded to the notion of anticipation, providing psychological evidence that anticipation plays a role in non-literal language processing. For instance, in Coulson's (2005) Space Structuring Model, a blending-based framework she uses to account for sarcasm, the sense of sarcasm arises out of the clash between the addressee's anticipation of the speaker's reaction and a counterfactual mental space. Similar views abound in the literature (e.g. Brône 2012; Coulson 2001; Coulson and Kutas 2001; Coulson et al. 1998; Davenport and Coulson 2011; Gibbs 2012; Giora 2003; St. George et al. 1994; Veale et al. 2013).

Linguistic typologists have further discovered that expressions coding information about anticipation have in some languages been grammaticalized into mirativity markers (DeLancey 1997; Aikhenvald 2012). A clear example provided by DeLancey to illustrate the idea comes from Hare, an Athabaskan language. Here the speaker utters the following upon seeing the addressee drinking.

(1) *ĩdõ* *lõ*

drink.2 MIR

"You're drinking!"

(DeLancey 1997: 40)

According to DeLancey's analysis, since the addressee certainly knows if they are drinking, the utterance in (1) lacks communicative force. What the mirative *lõ* marks is therefore the speaker's surprise at the fact that the addressee is drinking. The surprise is not related to the source of the new information, but is about the speaker being surprised by the new information.

Much like spoken language, music necessarily unfolds in time, a fact that makes it natural for composers to exploit the effects made possible by anticipation. Indeed, composers have long recognized the emotional potential of anticipation. One of the most well-known examples comes from Joseph Haydn's Symphony No. 94, popularly known as the "Surprise Symphony." In the second movement, the composer builds up a soothing, gentle melodic pattern, repeating it several times so that the listener becomes accustomed to it, before abruptly distorting the listener's anticipation with a note that is too high, too loud, and too different in timbre, surprising the audience (Levitin 2006). In the (once) discomforting "Augurs of Spring" of Igor Stravinsky's groundbreaking *Rite of Spring*, by accentuating upbeats and deaccentuating downbeats,

the composer intentionally decouples downbeats and accented beats, making the rhythm highly unpredictable and therefore producing a sense of uncertainty in the listener (Huron 2006). Another mastermind of manipulating the listener's anticipation in Western music is Richard Wagner. In *Tristan und Isolde*, one of his mature music dramas, the resolution for the dissonance of the renowned Tristan chord is deferred for almost four hours, a revolutionary compositional innovation that is integral to the music drama being so emotionally gripping (Bernstein 1976; Magee 2000).

Besides composers, who exploit the effect of anticipation knowingly or unknowingly, musicologists have likewise recognized the relation of anticipation as well as the effects it brings about (e.g. Levitin 2006; Meyer 1956). Approaching anticipation from experimental psychology, Tsai et al. (2014) present data of skin conductance, a physiological correlate of emotional experience, demonstrating anticipation's ability to impose emotional effects on the listener. A more comprehensive study comes from Huron (2006), where, based on prior empirical research, he proposes a framework for analyzing anticipation that distinguishes five psychological responses involved in human's coping mechanism for uncertainty, an ability he argues to be vital for survival from an evolutionary perspective. The five responses, taking place either before or after the outcome of an uncertain event, are imagination, tension, prediction, reaction, and appraisal. Though illustrated mainly through musical examples, Huron's theoretical framework is by no means confined to

music appreciation, but is aimed at accounting for the general mechanism with which humans cope with uncertainty, including situations where other modes of communication are involved.

Using the terms *anticipation* and *expectation* interchangeably, these past studies focus only on the temporal sequentiality between elements taking place at different points in time; that is, emphasis is only put on one's anticipation about future events based on past events and/or the present event. In (1), the speaker uses the mirativity marker because what they are seeing differs from what they anticipate based on their knowledge about the addressee as well as their encyclopedic knowledge. In the case of *Tristan und Isolde*, the listener likewise anticipates some sort of resolution based on the dissonance of the Tristan chord, the recognition of which is in turn rooted in the listener's frames (Fillmore 1985; Fillmore and Atkins 1992) for Western music.

In modes of communication where simultaneous events are possible, however, the relation between different events is not limited to being sequential; instead, a similar but distinct relation is also found between elements simultaneously represented in time. For instance, co-speech gesture or facial expression is not always congruous with the spoken language it accompanies. Such cases usually result from the speaker's reluctance to convey their intended meaning, whether intentionally or not (Wilcox and Shaffer 2006). Empirical studies have also shown that whether simultaneously represented multimodal inputs are congruous has an effect on the

understanding of the integrated message (Kelly et al. 2004; Wu and Coulson 2005, 2007). The prosody typically associated with ironic or sarcastic expressions (Attardo 2000; Attardo et al. 2003; Bolinger 1985, 1989; Kreuz and Roberts 1995; but see Sperber and Wilson 1981) can also be seen as the speaker's exploitation of the addressee's expecting the literal meaning of an expression to match the meaning communicated through its accompanying prosody.

Besides language, music is abundant with even more conspicuous examples. Take again Stravinsky's *Rite of Spring* for example. In the opening of the piece ("Augurs of Spring") the composer challenges the audience by juxtaposing chords of two different keys, making the music bitonal (Craft 1966; but see Chua 2007). This irritated the audience at the premiere and was partially responsible for the oft noted if not exaggerated riot in the audience, since at that time the listener would never expect chords of different tonalities to sound simultaneously[1]. In Salome's monologue, from the final scene of Richard Strauss's *Salome*, to stress the impossibility of Salome and Jokanaan's unification, the composer juxtaposes two chords—one in C sharp major (representing Salome), the other in C major (representing Jokanaan)—at the same

[1] While most of the audience was probably unaware of the bitonality of the music—one of the technical reasons why the music sounded discomforting to them—it still holds that part of the discomfort came from the unexpected temporal alignment of the simultaneously sounding notes. In addition to the music, the eccentric choreography was likely also responsible for the audience's reaction.

time. By simultaneously sounding two keys seemingly so adjacent and yet in reality so far apart in terms of harmony and on so many symbolic levels, the composer intentionally exploits the listener's expectation that simultaneously represented notes should be of the same key, imposing a sense of discomfort on the listener (Kramer 1990; Puffett 1989). Where language and music are both present, as in songs and operas, expectation can be exploited in an even more tangible manner. In the first act of Wolfgang Amadeus Mozart's opera *Don Giovanni*, the peasant maid Zerlina's words and the accompanying music bear contrasting semantics, suggesting the dishonesty and inner struggle of Zerlina (Hsu and Su 2014).

In films, where three or more modes of communication—the linguistic, the musical, and the visual—often coexist, incongruence between meanings communicated through different modes has also been exploited by filmmakers and analyzed in the film studies literature. Willemsen and Kiss (2013), for instance, look into the torture scene in Quentin Tarantino's black-comedy crime film *Reservoir Dogs*, in which Mr. Blonde uses a razor blade to ruthlessly cut off the right ear of a police officer who is held hostage. Accompanying this horrific scene, however, is not underscore, but Stealers Wheel's upbeat popular tune *Stuck in the Middle with You*, whose lightness is in stark contrast with what is taking place visually, amplifying the suspense. Bashwiner (2013) likewise examines the nuances of the cross-modal pairing of King George VI's speech and the Allegretto from Beethoven's Seventh Symphony in the film *The King's Speech*.

Experimental studies on the relation between simultaneously represented audio and visual inputs also abound in the literature (e.g. Bolivar et al. 1994; Boltz 2004; Boltz et al. 1991; Marshall and Cohen 1988; Shevy 2007; Vitouch 2001).

To better distinguish the two relations reviewed above, the term *anticipation* will be used in the present study to refer to the relation holding between temporally sequential events, whereas *expectation* will be employed where the relation is found between simultaneously represented events[2]. A clearer, visual representation of the two relations of anticipation and expectation is shown in Figure 1. Two modes of communication, X and Y, are depicted here, although there can certainly be more than two modes. X1, X2, and X3 are constructions in mode X (see Section 3.1.4; for now, suffice it to say that events are made up of constructions) taking place at different points in time, which are in turn represented by t1, t2, and t3. Likewise, Y1, Y2, and Y3 are constructions in mode Y at different points on the horizontal axis representing time. The temporally sequential relation of anticipation holds by definition between constructions that happen at different points in time and that have a sequential relation

[2] While "expectation" and "anticipation" are in most contexts used interchangeably in the English language and might therefore not be the perfect labels for the two relations in question, they are to my knowledge the best labels readily available. Terms often employed in relevant discussions, such as "congruence," "incongruity," and the like, do not really capture the distinction between simultaneity and sequentiality, as those terms denote phenomena that can be found in both simultaneous and sequential relations.

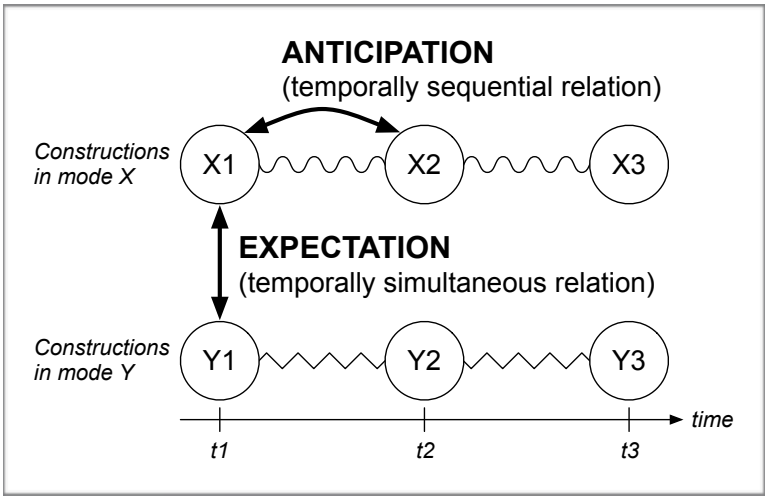


Fig. 1. Expectation and anticipation.

to one another. In other words, it is a relation parallel to the progression of time. Although the Figure only marks the anticipation that holds between X1 and X2, the same relation can be found between any other constructions that are temporally sequential to one another (e.g. between Y1 and Y3), including even constructions of different modes (e.g. between Y2 and X3). Expectation, the temporally simultaneous relation, on the other hand, is shown in Figure 1 as holding between X1 and Y1, both of which take place at t1. The relation of expectation likewise exists between other simultaneously represented constructions, such as X2 and Y2, X3 and Y3, and so on. For certain modes of communication—such as music—where more than one construction can be simultaneously represented—such as juxtaposed chords—the relation of expectation can also be found within the same mode, between constructions taking place at the same time. As seen in the Figure, it is a relation perpendicular to the progression of time.

While the idea of sequentiality has long been a core concern in linguistics, the notion of simultaneity has yet to receive much attention, probably owing to the common practice in linguistic analysis of reducing language to a purely monomodal means of communication, disregarding the abundance of information communicated paralinguistically—that is, through other co-speech modes (but see Bolinger 1965). On the periphery of traditional linguistic analysis, signed language and gesture studies, on the other hand, with a focus on the visual-spatial modality, have recognized the importance of simultaneity (e.g. Vermeerbergen et al. 2007). As pointed out by Perniss (2007), the existence of multiple, independent visual articulators avails signed languages of the possibility of representing simultaneously multiple independent meaningful units, giving rise to constructions that are peculiar to signed languages.

Identifying the temporal notion of simultaneity in communication, these signed language and gesture studies nonetheless still leave unaccounted for how simultaneously represented events are related to one another. This very missing connection, I argue, is precisely expectation, defined above as the relation that holds between simultaneously represented constructions. It is the mismatch or even clash between simultaneously represented elements that prompts the effect found in such examples as the bitonality of *Rite of Spring*, the text-music incongruity in *Don Giovanni*, and the mismatch between Mime's words and leitmotif mentioned in Chapter 1. Though ubiquitous in our daily life and indispensable for our understanding of the

most ordinary, the relation of expectation has hardly been recognized let alone analyzed in the literature, except cursory mentions in some recent studies (Hsu and Su 2014; Veale et al. 2013), where the idea is only alluded to in passing. In the following chapters, therefore, having recognized the relation of expectation and distinguished it from anticipation, I seek to further explore the role expectation plays in human cognition. More specifically, I will demonstrate what cognitive mechanisms are involved in the processing of expectation, further providing a possible set of analytical tools for studying cases of multimodal communication where expectation is involved.

The meticulous reader might have noticed that there are two kinds of anticipation. There are occasions on which we were not anticipating anything in particular but still get surprised anyway. For instance, a student going into their professor's office during the professor's office hours is certainly to be surprised if they find the professor gaming on the computer. The student may not have been consciously anticipating anything specific beforehand, but they did have a general idea, consciously or subconsciously, of what they should anticipate seeing the professor doing. There are also occasions where one has in mind some fairly specific anticipation, which is later violated. This is probably a more prototypical case when people talk about anticipation. For example, an experienced concertgoer may be extremely surprised when, say, the oboe enters ten bars early. That is, the concertgoer is surprised because they are anticipating the oboe only at a precise moment. These two different kinds of

anticipation roughly correspond to Reisenzein et al.'s (2012) two degrees of schema discrepancy: one where the reality is incongruous not with some explicit anticipation but with one's general beliefs, and one where the reality clashes with one's specific, pre-existing, and oftentimes explicit beliefs. In his study on anticipation in music, Huron (2006) likewise distinguishes between two kinds of anticipation: schematic and veridical. While schematic anticipation is not explicit and is based on the listener's prior schematic knowledge, veridical anticipation results from the listener's experience with prior tokens of the same event. As different as these two kinds of anticipation may seem on the surface, they are in fact two points on a single continuum, at least on the level of psychology. Whereas in the latter case it is clear the concertgoer is anticipating something very concrete and specific, one cannot really say that the student in the former case is not anticipating anything at all. As intangible as it may be, the student must have a general if not rough idea of what they are going to encounter. Accordingly, the distinction between the anticipation in the former case and that in the latter will be disregarded in the present study, since their difference is more of a matter of degree than category.

Another issue that is noteworthy is that the ideas of simultaneity and sequentiality exist on two levels: perception and cognition. Take the visual modality for example. When we look at an advertisement with two juxtaposed images, the focus of our eyes can only be on one of the images at a time. That is, we need to look at one of

the images first before we can focus on the other. We have the freedom of choosing what to look at first, how much time to spend on each of the images, if we want to look at one of the images again, etc. The progression of time is not an inherent formal property of the juxtaposed images. It certainly still takes time to perceive the images, but the way they are perceived is not set in time, but determined externally. The viewer of the images can decide how their visual perception unfolds in time, on which the images themselves have no say. In cognition, on the other hand, since it is one advertisement with two juxtaposed images, we are prompted to process them together —we compare the possible meanings of the two images and try to make sense of the ad as a whole (see Section 3.1.5). In short, in modalities like the visual, due to the nature of our sense organs (e.g. our eyes' ability to focus on only one thing at a time), simultaneity in perception is not possible, even though messages perceived sequentially can be processed simultaneously in cognition.

On the contrary, in modalities like the auditory modality, we are not given the option to choose what to hear. As long as there is any sound wave perceptible to the human ear around us, we hear the sound. If there is more than one sound wave around us, we have no choice but to perceive all of them at once. We are not allowed, say, to listen only to the lyrics of a song *and then* to the music. We do not get to choose the pace at which we perceive sounds. We do not get to rewind a certain sound and listen to it again. We hear whatever we are exposed to, in whatever way predetermined by the

source. We could, of course, listen only to the music to a song without paying attention to the words or without understanding the words, but even in those cases we are still forced to perceive both linguistic and musical inputs in the way predetermined by the song. Simultaneity in perception in modalities like this is therefore not just possible but in many cases imposed upon us. Simultaneity in cognition, which is not constrained by the nature of perception, is of course also possible for these sorts of modalities. We can easily process information from different channels simultaneously.

In short, while simultaneity in cognition is possible regardless of the modes of communication, simultaneity in perception is in some modalities constrained by the nature of our sense organs. Given the significance of the distinction, modes of communication where simultaneity in perception is possible and those where it is not deserve separate investigations and should not be conflated (see Arnheim 1974 for a more detailed discussion on this distinction). In view of this, the present study will focus only on modalities where the progression of time is an inherent formal property; that is, modalities where simultaneity in both perception and cognition is possible.

## 3. A COGNITIVE TOOLBOX FOR EXPECTATION.

Having identified the existence of the simultaneous relation of expectation, we now need to delve deeper to have a finer-grained view of the nature of the relation, so as to determine what cognitive operations are at work when people process expectation; that is, how human cognition handles and makes sense of expectation. Although expectation is a hitherto overlooked phenomenon, insight from related research can still be drawn upon and serve as an analogy.

In musicology and multimedia studies, theoretical frameworks have been put forward to account for cases in which inputs from different modalities interact. Contemplating on the relation between language and music, Cook (1998) postulates three possible models: conformance, complementation, and contest. The conformance model applies where the language and music are consistent instead of coherent (in Lakoff and Johnson's [1980] sense). In cases in which the language and music are coherent, the complementation model applies where the two modalities are contrary to each other, and the contest model where the two are contradictory. In researching the interactions between visual and musical meaning, Marshall and Cohen (1988) and Cohen (2010, 2013) propose the Congruence-Association Model. According to the Model, information associated with music that does not overlap with the information communicated through the film it accompanies has an effect on the cognizer's interpretation of the visual information. As Cohen (2010: 886) puts it, "From moment

to moment, the audience member extracts information from non-diegetic sources to generate the emotional information he or she needs to make a coherent story in the diegesis."

While these proposals provide valuable insight into multimodal meaning construction, they are nonetheless overly coarse-grained. Cook's three models of multimedia, for instance, merely identify three discrete categories of language-music combinations, overlooking the fact that cross-modal relations are often highly nuanced and seldom reducible to a series of binary parameters. The Congruence-Association Model, on the other hand, while accounting for music's impact on the film it accompanies, neglects the multidirectional nature of cross-modal meaning interactions.

Though not strictly linguistic or multimodal, Teng and Sun's (2002) study on image alignment may also serve as a starting point. According to their Image Grouping Hypothesis, humans tend to see pictorial elements that are symmetrically aligned in space as belonging to the same category. Accordingly, when juxtaposed images are originally from different categories, the alignment invites the viewer to think of them as belonging to the same category. Depending on the degree of difference between the images, a pictorial grouping, pictorial simile, or pictorial oxymoron is then formed. If the juxtaposed images are congruous and of the same category, a pictorial grouping will be in place; if the aligned images come from different categories, the alignment will prompt the viewer to see the images as belonging to the same category, resulting in a

pictorial simile; if the aligned images are neither of the same category nor congruous, the viewer will still be inclined to view them as coming from the same category, and a pictorial oxymoron will be formed, further eliciting novel and open-ended interpretations from the viewer.

*Eo ipso*, humans are also inclined to view elements that are "temporally aligned" as conveying the same meaning. When listening to a song, for instance, the listener would usually expect to hear words and music that share similar meaning; that is, the simultaneous representation invites the listener to expect the linguistic and musical semantics to match or complement each other. When this is not the case, as in the *Siegfried* passage, where the music accompanying Mime's words is unexpected, the temporal alignment still invites the listener to view the incongruous language and music as being intended by the composer to convey a joint, often emergent meaning, which is usually much richer than in cases where everything is within expectation.

Albeit cogent and insightful, an analogy from a related phenomenon is, after all, an analogy at most. To tackle the peculiarities of expectation more directly, and to account for the dynamic multimodal meaning interactions therein, in the following I consider and review several theoretical notions which are relevant to the understanding of expectation, but which have been lacking in existing theoretical frameworks. Also examined are more familiar notions that deserve recapitulation, given their pertinence to the understanding of expectation.

## 3.1. PREPARING THE TOOLS.

Despite the fact that the present study started out with an opera passage, and that due to the limitations of the scope only examples involving language and music will be examined (see Chapter 4), the following theoretical tools are not limited to analysis of language and music, but are general to the channels of communication available to humans, which also include for instance gesture, body movement, eye gaze, and even smell. This is based on the core assumption shared among cognitive linguists that the cognitive resources affording humans the ability to use language are not autonomous, but are shared by our other cognitive capabilities, such as music appreciation, logical reasoning, emotions, spatial perception, and so forth (see for example Croft and Cruse 2004; Geeraerts and Cuyckens 2007). Although at the current stage the extension to modalities other than language and music awaits further investigation, the following attempt is hoped to be as general to human cognition as possible.

Given that, in many cases of multimodal communication, meaning is communicated through different channels of perception—that is, the modes involved are often also "multi-perceptional"—the conventional, monomodal labels such as "the listener," "the hearer," "the viewer," etc. do not really suffice. The term "the cognizer" is therefore adopted to refer to the person undertaking the cognitive task of making

sense of the inputs perceived, whether consciously or unconsciously, regardless of the channels through which the inputs are perceived.

### **3.1.1.** TIME.

As suggested by the main contrast between anticipation and expectation in Figure 1, the dimensions on which sequentiality and simultaneity respectively lie mark some of the most fundamental parameters in expectation research. Starting from sequentiality, the axis along which the relation of anticipation holds is the progression of time, a physical property that is (on the human scale) linear, unidirectional, and unstoppable. Though crucial not just to our understanding of expectation but indeed to many other everyday cognitive activities, the notion of the progression of time happens to be the one thing that is missing in most existing theoretical models, if not all. Although most cognitively oriented models are claimed to be able to capture the dynamicity of meaning construction, few, if any, explicitly incorporate the progression of time as an actual variable. Instead, most models only offer synchronic sections of meaning construction processes, which by definition span over a period of time. That is, most existing cognitively informed models recognize the dynamic nature of their subject matter, but few, if any, are actually capable of capturing the dynamicity, let alone properly analyzing it. The significance of dynamicity in meaning construction makes the progression of time a variable essential to any relevant theoretical

framework. By explicitly incorporating the progression of time into the analytical framework, we can capture and examine not only synchronic sections, but also the relations between them.

### **3.1.2.** CONCEPTUAL DISTANCE.

The axis along which the simultaneous relation of expectation exists, on the other hand, is the conceptual distance that holds between events taking place at the same time. The notion of conceptual distance is crucial in that the relation between inputs represented in simultaneity cannot be reduced to the dichotomy of congruity and incongruity. Indeed, as Kelly et al. (2010) have observed based on empirical research on co-speech gesture, the extent to which speech and gesture are incongruous to each other is inversely correlated with the accuracy of the cognizer's comprehension of the multimodal inputs. Much like the progression of time, the notion of conceptual distance also happens to be missing in most existing accounts. Whereas relevant topics such as cognitive dissonance have received more thorough consideration in psychology and sociology (e.g. Festinger 1957; Van Overwalle and Jordens 2002), discussion in linguistics on mismatch, incongruity, or dissonance is rather limited, save for occasional mentions in studies on humor, irony, and sarcasm (e.g. Attardo 1997; Attardo et al. 2002; or Brône 2012 for a review), most of which merely identify the

existence of dissonance or incongruity, overlooking what precisely it is that differentiates them from cases where there is no incongruity.

Since concepts are not entities in a physical space, conceptual distances—distances between concepts—by definition can only be metaphorical. It follows that there is no way to really measure conceptual distances, which do not really exist. What needs to be recognized here is that by conceptual distances we are really talking about degrees of similarity or dissimilarity between concepts, and that conceptual distances can be measured only indirectly through measuring degrees of similarity between concepts. Moreover, levels of similarity can themselves be indirect measurements in the sense that properties of most concepts fall on more than one dimension. Adding to the complexity of the issue is the fact that not all dimensions are scalar, and that not all dimensions have both positive and negative values.

In order to measure and compare conceptual distances, therefore, we need to identify the dimensions on which the properties of the concepts in question lie, as well as the nature of these dimensions. Once we have identified the dimensions and their qualities, we will then be able to construct a conceptual space in which metaphorical distances can be captured and measured. Like in geometry, the conceptual space does not need to be one humans are able to perceive; that is, a three-dimensional space. Conceptual distances in a conceptual space can be calculated even when the number of dimensions well exceeds three. The fact that conceptual distances are themselves one-

dimensional does not mean differences among concepts are one-dimensional, however. In cases where common dimensions exist among the concepts compared, differences on those dimensions can be easily measured. Where there is no common dimension, conceptual distance is the only possible measure, but this does not mean the differences are one-dimensional; the differences simply cannot be directly compared.

Since describing differences between concepts in terms of physical distances is metaphorical *per se*, it follows that there is no physical distance to measure, and that it is impossible to measure the real, absolute distance between any two concepts. However, it is possible to measure the metaphorical distances between concepts, which could help us understand the degree of similarity or dissimilarity between different concepts. Various attempts have been proposed in the literature as to how conceptual distances can be compared, measured, or operationalized. Relatively familiar to linguists, the semantic maps model (e.g. Croft 2003; Kemmer 1993; Stassen 1997) proposes a way to identify the relative distances between different concepts and whether two concepts are contiguous, without attempting to capture more precise distances. Another notable attempt comes from Gärdenfors (2000, 2014), who proposes a way to conceptualize concepts in our brains as a geometric space. According to Gärdenfors, any given idea is situated in a conceptual space made up of one or more

dimensions and has values on these dimensions.[3] More specifically, the idea of domain, in its conventional sense in the linguistic literature, can be thought of as a set of integral (as opposed to separable, in Gärdenfors's [2014] sense) dimensions, where concepts in a single domain have values on all those integral dimensions. A property, then, is defined as a region in a conceptual space. Also noteworthy is that, in Gärdenfors's model, what matters are not the referents or the representations, but the inter-relations of the referents and those of the representations.

As for measurement of distance, among the various multifactorial analysis methods proposed, one that is widely used is multidimensional scaling (henceforth MDS). In MDS, data of similarity, whether from empirical evidence or subject judgment, is analyzed by the computer, which will then identify dimensions along which correlations can be identified. Admittedly, it is often unclear precisely what those dimensions identified by the computer represent, but it is probably the most convenient and operationalizable way currently available to measure, or at least compare, conceptual distances between different concepts.

---

[3] Gärdenfors's (2000: 5) claim that there is a "tight connection" between conceptual distances and similarity judgments is in fact misleading. Conceptual distance and similarity are not two independent concepts; conceptual distance is only a means through which we understand similarity metaphorically.

Another notion of conceptual distance that the present research can draw upon is the one proposed by the linguistic typologist John Haiman. In his analysis of iconicity, Haiman (1983) argues that the formal distance between syntagmatic linguistic elements is motivated by the distance between the concepts that the linguistic elements represent. That is, the "closer" two concepts are to each other, the closer their respective linguistic forms are syntagmatically, and vice versa. While Haiman's analysis is about how the conceptual distance between two concepts is reflected in their distance in form, which is a distance between temporally syntagmatic events, and which is therefore a distance parallel to the progression of time, in cases where simultaneous events are possible, conceptual distances also exist in between any simultaneous (that is, temporally paradigmatic) events, except that these are distances perpendicular to the progression of time.

Consider again Figure 1. Based on Haiman's view, X2's being temporally closer to X1 than X3 is iconic, considering the fact that the semantics of X2 is closer in conceptual space to X1 than X3 is to X1. In other words, what Haiman looks at is how conceptual distances are reflected in the sequentiality of strings of forms. Since conceptual distances are supposed to be independent of the progression of time—that is, the times at which two concepts (NB: not two forms) are evoked have nothing to do with the distance between them in conceptual space—it appears reasonable that we look at conceptual distances between events taking place not just in sequentiality but

also in simultaneity. One should therefore consider not just the conceptual distance between X1 and X2 or Y2 and Y3, but also those between X1 and Y1, X2 and Y2, X3 and Y3, and so on.

In addition to Gärdenfors's and Haiman's definitions of conceptual distance, it is worth mentioning that, in cases like Figure 1, while X and Y are two distinct modes, concepts evoked through forms in those two modes belong to the same conceptual space. That is, a concept's location in a conceptual space remains constant regardless of the mode in which it is communicated. Equally noteworthy is the idea that, even in cases where there is no cognitive dissonance, conceptual distances between concepts still oscillate from one point in time to the next; the changes are simply not as noticeable as those in cases involving incongruity.

The idea of the conceptual distance between simultaneous events is crucial in any model that attempts to account for multimodal meaning construction. By identifying the conceptual distances between different pairs of events, one would furthermore be able to capture the dynamic change in the distances, therefore comparing them, in turn seeing how they contribute to the dynamic meaningfulness of the global structure. In the present study, the idea of conceptual distance will be adopted in Gärdenfors's (2010, 2014) sense, with Haiman's definition as an anchor in linguistics. Empirical measuring of conceptual distance, however, is beyond the scope

of the current study, which aims to serve as a theoretical starting point on which further empirical studies can be based.

### **3.1.3.** Conceptual integration and disintegration.

Having identified the dimensions along which different events in different modes can be located, the next step is to see how these events can interact with one another. With the capability to handle multiple, simultaneously represented inputs and the ability to account for the emergence of novel meaning, the Conceptual Blending Theory put forward by Fauconnier and Turner (2002) provides an optimal theoretical tool. Indeed, the framework of conceptual blending has proven to be insightful for studies on not just language but also multimodal phenomena (e.g. Zbikowski 1999, 2002; Hsu and Su 2014).

Proposed against the background of the Conceptual Metaphor Theory (Lakoff and Johnson 1980), the Conceptual Blending Theory aims at providing a comprehensive account of a cognitive capability that defines the human being and that is general to a plethora of cognitive tasks humans are capable of—conceptual blending. Essentially, conceptual blending is a process in which two or more inputs that are incompatible in real life get creatively integrated, yielding blended meaning that is absent in the original inputs. The three processes of composition, completion, and elaboration, which are involved in conceptual integration and meaning emergence, are

further identified by Fauconnier and Turner (2002): In composition, elements and relations from input spaces are selectively recruited and projected into a blended space; in completion, the cognizer's contextual, encyclopedic, and schematic knowledge is added to the blend; in elaboration, the elements and relations recruited in the previous processes are simulated in the fictive scenario of the blend, creating imaginative interpretations and rich emergent meaning.

An example that illustrates the basics of the ubiquitous process of conceptual blending is the Buddhist Monk Riddle (Fauconnier and Turner 2002): Essentially, a Buddhist Monk spends a day walking up a mountain. After a few days of meditation on the top of the mountain, he spends another day walking down the mountain via the same path. The riddle goes: Is there a point of the path which the monk occupies in the same hour on his two journeys? To solve this riddle, the cognizer needs to construct an input mental space in which the monk walks up the mountain and a second input space in which the monk goes down the mountain, with details such as time, location, and path in each of these input spaces. Elements and relations in these two input spaces are then selectively projected into a blended space, in which a monk is walking up the mountain and a second monk is walking down the mountain. Through the impossible integration of the two incompatible inputs—incompatible in terms of time, identity, and so on—the meaning of the two monks' impossible encounter emerges, allowing the cognizer to decipher the riddle. As this example shows, even the most

commonplace of our daily life requires such complex and extraordinary cognitive processing.

Importantly, as Fauconnier and Turner (2002: 119) point out, "Integration and compression are one side of the coin; disintegration and decompression are the other." Conceptual blending also includes a process in which part of the emergent meaning gets projected back into the inputs, modifying the cognizer's initial understanding of the inputs. This is what Fauconnier and Turner (2002) call disintegration, decompression, reverse mapping, and the like. The inconsistency in their naming, as well as the lack of a clear definition of the process, has been criticized by Hougaard (2002, 2005) and Bache (2005), among others, who advocate the centrality of conceptual disintegration in conceptual blending, a model that aims at being all-encompassing but that only focuses on integration. In the present study, the term disintegration (instead of decompression or reverse mapping) will be used to refer to the process whereby emergent elements and relations in the blended space are selectively projected back to the original inputs.

The fact that perception and cognitive processing unfold in time suggests that, where conceptual integration is involved, conceptual disintegration is also at work. The role of conceptual disintegration is even more indispensable in cases where incongruity exists in between inputs. In the aforementioned *Siegfried* passage—a usage event that unfolds over time and contains incongruity—it is hardly possible for the listener to

realize the dishonesty of Mime's words the moment the passage starts. Rather, it has to wait until some later point in time when the listener has heard enough text and music to be able to make sense of the cross-modal mismatch and therefore to modify their initial understanding of Mime's language-music combination prior to that point in time. In fact, conceptual disintegration is at work even where no mismatch is present. It is hard, if not impossible, to find a case where the cognizer's understanding of an event is not later modified in accordance with their understanding of subsequent events. The ubiquity and relevance of conceptual disintegration in dynamic meaning construction makes it a necessary complement to conceptual integration.

Equally crucial to conceptual blending is the identification of vital relations (Fauconnier and Turner 2002), which hold between elements from different inputs, as well as between elements inside the blended space. Vital relations define the way input spaces are related to one another. It is precisely because of the existence of vital relations between the input spaces of a conceptual integration network that the cognizer attempts to blend the inputs in the first place. In the Buddhist Monk Riddle, the vital relations between the two inputs include among others Identity, since the monk in the first input space is the same person as that in the second, and Time, as the second input takes place subsequent to the first. While the existence of these vital relations invites the cognizer to blend the two input spaces, the identification of the vital relations guides the cognizer in coming up with the emergent structure in the

blended space, in which the vital relations are compressed into Uniqueness between the two monks.

In cases where simultaneous events are possible, among the numerous basic vital relations that can be found between inputs, the defining vital relation is Expectation. Recall the implications drawn earlier from Teng and Sun's (2002) study. It is precisely the expectation that simultaneously represented events should "support" one another that invites the cognizer to view them as belonging to the same category. Where incongruity exists between inputs, without this expectation the inputs would not even be compared, let alone blended. Take the *Siegfried* passage for example. If it were not for the expectation resulting from the temporal alignment, the listener would have no reason to view Mime's words and the leitmotif for Mime's complaint as jointly constructing some sort of emergent meaning. Besides Expectation, a vital relation often found between temporally sequential inputs is Anticipation, as the cognizer would usually anticipate future events based on past or present events.

### **3.1.4.** CONSTRUCTION.

Linguists have long been arguing over what the basic unit, or building block, of language is, as an agreement on this issue is the first step to comparisons within and across languages. The necessity of identifying the basic unit of analysis is even more pressing when it comes to multimodal studies, where researchers need to not only look

at the modalities involved but also compare the modalities and examine the cross-modal relations. It is not enough, for example, to analyze only the text and the music of a song separately, without paying any attention to the intermodal dynamics between the text and the music. In other words, just as compositionality in the strict sense seldom holds for more complex structures in language, the meaning of language-music combinations is hardly ever the sum of its linguistic meaning and musical meaning; rather, a great part of the meaning often only emerges when the cognizer compares the meaning across the two modes.

Nevertheless, despite various attempts from numerous scholars, the preferred practice in recent decades of breaking complex phenomena down into smaller, more manageable, and preferably recursive parts seems to have made a consensus on the definition of the basic unit in language unlikely, a debacle probably attributable to the inconvenient fact that human languages are messier and less systematic than linguists would like. Digressing from prior attempts' sole attention on linguistic form and shifting the focus onto both linguistic form and function, construction grammar, a strand of inquiry that recognizes the centrality of form-function relations in language, appears to be capable of breaking through the bottleneck and offering an account of language structure that is potentially more coherent.

Consisting of several theoretical proposals, construction grammar is in fact a cover term for a conglomerate of various approaches to linguistic grammar (e.g.

Fillmore et al. 1988; Goldberg 1995, 2006; Croft 2001; Bergen and Chang 2013; Michaelis 2013; Steels 2013), including those whose proponents do not explicitly identify themselves as construction grammarians (e.g. Langacker 2008; Tomasello 2003). Among the various takes on construction grammar, proposals from Goldberg (1995, 2006), Croft (2001), and Langacker (1987, 2008) converge to a large extent in their definitions of constructions, which they see as the basic unit, or building block, of human languages. For them, a construction is essentially a pairing of form and function, where the form is not constrained by the traditional word-based parts-of-speech view of linguistics, and where the function includes not just semantic meaning in the narrow sense but also pragmatic inference, contextual meaning, grammatical function, etc. In other words, a form-function pairing is a construction however unconventional the scope of the form is, and however "pragmatic" the function is. For instance, though not a "word" in the conventional sense, the English plural marker *-s* is a construction, since it takes form and has the function of marking plurality. The phrase *to rub someone the wrong way* is also a construction, despite the fact that it consists of multiple "words," since the meaning of the phrase is associated with the entire phrase and is not compositional. By the same token, other morphemes, phrases, idioms, and certainly "words" in the conventional sense can all be constructions, as long as they are pairings of form and function. Detached from the orthographic form, the approach construction grammarians take on linguistic grammar provides a more

coherent way to look at language, as it pays equal attention to both form and function, the latter of which can be located in the supposedly universal conceptual space (Croft 2001; cf. Section 3.1.2).

The functionally oriented view shared by (some) construction grammarians that form-function pairings—instead of "words" or parts of speech—are the building blocks of language makes it valid not only crosslinguistically (Croft 2001) but possibly also across the modes of meaning communication available to the human being. Indeed, as Langacker (2008: 457–8) suggests, the possible forms a construction can take include "the full phonetic detail of an utterance, as well as any other kinds of signals, such as gestures and body language (conceivably even pheromones)." Whatever the modality, the purpose of communication is to get meaning across through some sort of form. This is also the case in music, which is arguably made up of constructions as well, as all audio stimuli in music serve some sort of function. While the form can be a motif, a chord, a progression of harmony, or even just a note[4], the function can be, say, stimulating emotional affect, creating association with ideas or object, signaling a cadence, or delaying resolution. For example, the leitmotif representing Mime's

[4] For instance, the long-held E flat note played by double-basses in the very beginning of Richard Wagner's *Das Rheingold* is said to symbolize the origin of the world (Newman 1949), from which everything evolves, and to which everything eventually returns. The recurring hammer blow in the Finale of Gustav Mahler's Sixth Symphony, which repeatedly shatters all remaining hopes (Monahan 2007), is also a case in point.

complaint mentioned in Chapter 1 is clearly a construction, since it is an idiosyncratic pairing of form (see Section 4.1.1 for the actual form) and function (the association with Mime's complaint). The tonalities of C sharp major and C major in Richard Strauss's *Salome* are likewise constructions, as they are associated with the characters of Salome and Jokanaan, respectively, despite the fact that tonalities—the forms of these constructions—are defined on a more schematic level and can be realized in a variety of ways.

Drawing upon examples from popular and Western Classical music, in recent years musicologists have also started to recognize the theoretical potential of construction grammar as a way of analyzing the structuring of musical forms and functions (e.g. Gjerdingen and Bourne 2015; Zbikowski 2011). In particular, recognizing the parallels between language and music, Zbikowski (2011: 92) defines musical constructions as "sequences of musical materials [...] that serve as sonic analogues for dynamic processes," a view that echoes the tenets of construction grammar, even though he also stresses at the same time that "the different functions [language and music] serve within human culture suggests that their grammars will also be different" (Zbikowski 2011: 92).

Intended as a radical reexamination of linguistic grammar, construction grammar also tackles traditional issues in syntax. As Langacker (2008) lays out specifically, constructions can themselves be composite structures, encompassing

component constructions. The relationship between a composite construction and its component constructions need not be compositional, which is precisely why complex forms should also be seen as constructions on their own with their idiosyncratic pairings of forms and functions. A similar view is also shared by Zbikowski (2011), for whom the syntactic processes that organize basic musical constructions into more complex structures are also constructions. Examples of composite structures include the English idiom *kick the bucket*. While the idiom is clearly a construction with a highly idiosyncratic meaning, *kick* and *the bucket* are also form-function pairings themselves. Similarly, while the aforementioned juxtaposition of Salome's and Jokanaan's chords in Richard Strauss's *Salome* is a composite construction with the function of highlighting the incompatibility of the union, Salome's and Jokanaan's chords are still constructions in their own right, also contributing to the meaning of the composite construction, which has emergent meaning absent in its component constructions.

Depending on the cognizer's familiarity with the subject matter, constructions may differ even across cognizers of the same community, a fact that is possibly even more common when it comes to musical constructions, as musical meaning is for most cognizers less tangible than linguistic meaning. The present study does not aim to provide an operationalizable definition of musical constructions so much as to offer a possibility of viewing language and music in a functionally justified way that is both intra- and inter-modally coherent.

### 3.1.5. INTENTION.

Whenever people communicate, and in whatever mode, they can be said to communicate for a reason—to change the addressee's knowledge structure (cf. Clark 1996; Grice 1957). For instance, when the speaker shares some gossip with the addressee about some common friend, an addition or modification is made to the addressee's assessment of that common friend (and possibly of the speaker). When the speaker makes chitchat with the addressee, however shallow the content is, the latter's understanding of the former becomes more thorough and comprehensive. For the speaker, to make changes in the addressee's current knowledge structure is a major component of the speaker's intention in using language. The speaker's intention is of course not limited to the content and purpose of their actual speech—to accomplish their goal, the speaker would sometimes choose not to reveal all pertinent information —but the speaker's actual speech is always part of their intention, and is one of the very few clues available to the addressee about the speaker's intention.

The topic of intention has long intrigued and puzzled people interested in language, including philosophers, cognitive scientists, and, of course, linguists. Placing major emphasis on function, the issue of intention has been central to functional linguistics almost since the onset of the field. As Chafe (1977a, b) and Croft (2007) argue, for instance, communication through language is a process in which the speaker encodes their intention in language, and in which the addressee decodes the speaker's

intention based on the linguistic input from the speaker. In this way, communication can even be seen as an embodied form of intention.

As the present study is concerned with how the cognizer is able to make sense of meaning emergent from congruous or incongruous information communicated simultaneously through different modes, the intention of the person communicating this meaning is of central importance to the analysis. Recall the *Siegfried* passage in which Mime sings words of love and care to the leitmotif of complaint. The listener needs to make sense of the incongruous, seemingly mismatched pairing of language and music in order to understand what Mime really means, to grasp what the intention of Mime's character really is—namely that he is pretending not to hold a grudge against Siegfried as part of his master plan to seize the Ring. On top of this, by rationalizing the language-music mismatch, the listener also understands the composer's intention in composing the passage in this way, which is to highlight Mime's dishonesty. In other words, the listener is able to understand not just the fact that Mime is holding a grudge (the character's true intention), but also that Mime is a dishonest being who would hide his true intention for his personal interests (the information the composer intends for the listener to grasp).

That said, no participant in a communication event has direct access to the other participants' intentions. As the semiotician Roland Barthes (1964/1977) pointed out over half a century ago, the author of a text dies the moment the text is finished. To

wit, the author does not have authority over the interpretation of their own text, and it is the reader's instead of the author's interpretation of the text that matters. In fact, even if the author does have a say on the interpretation of the text, the author's intention is still inaccessible. Chafe's (1977a, b) and Croft's (2007) models of verbalization clearly capture the fact that the speaker's intention has been "translated" multiple times before the addressee even hears the words, and even more times before the addressee understands the speaker's intention. The speaker's intention and the addressee's interpretation of the speaker's intention are never identical. Interpersonal communication is successful not when the addressee has fully grasped the speaker's intention, but when there is enough overlap between the speaker's intention and the addressee's interpretation of it (Langacker 2008). By the same token, the listener of the *Siegfried* passage has direct access neither to the composer's nor to Mime's character's intention, and neither "exists" so as to have tangible bearing on the listener's interpretation. In light of this, whenever "the composer's intention," "the character's true intention," or the like is mentioned in the present study, the term refers to the cognizer's understanding of the intention in question, as that is what really matters in communication. Furthermore, while the plural form of the word *intention* can be used to refer to one's general purpose or attitude toward someone or something—as in *my intentions are good, that's a guy with bad intentions*, etc.—the word will only be used to mean one's will and volition in the narrow sense in the present research. In the

following, therefore, only the singular form will be used even where the collocation pattern favors the plural.

The issue of intention does not stop at the identification and delimitation of it. The speaker's intention, for instance, can often be broken down into multiple levels[5] simultaneously, with only some of the levels exposed to the addressee. A politician can make the meaning they intend for their addressees to understand very different from their true, unspoken intention. As the *Siegfried* passage also exemplifies, the meaning Mime intends for Siegfried to believe in is almost the opposite of Mime's real, hidden intention. In this case, while the listener has access to both levels of Mime's intention, without the assistance of leitmotifs Siegfried's character can only understand the meaning of Mime's literal words. Whether in theater or in real life, such phenomena not only prevail in communication, but can often be even more convoluted. However, since the focus of the present study is a cognitive one with primary emphasis on the cognitive processing of the cognizer in dealing with multimodal cases of meaning communication, the finer-grained details of intention, which really are more pertinent to the social-interactional aspect of communication, will not be incorporated into the main analysis. Issues pertaining to intention will be addressed after the primarily cognitively oriented discussion (see Chapter 5).

---

[5] Here the term "level" is used as a generic term, and is not to be confused with Clark's (1996) specialized use of the word.

Another notion pertaining to intention that is particularly crucial to cases of multimodal communication where cross-modal mismatch is observed is the intentionality of alignment. In the typical conversation, the speaker's speech, gesture, eye gaze, posture, etc. are intended to be aligned as an integrated communicative act. In songs and operas, language and music are temporally aligned by the composer to convey their intended meaning and achieve their intended effect. In their research on gesture-speech relations, Kelly et al. (2007) lend empirical support to the fact that the addressee's knowledge of whether the gesture-speech combination is intentional matters. That is, the addressee would be more likely to try to make sense of a gesture-speech combination if they know that the gesture and the speech are related and meant to be interpreted together. By the same token, when listening to a song, however conceptually dissonant the language and the music are, one can assume that the listener would attempt to make sense of the language-music combination, a fact precisely attributable to the listener's assumption that the multimodal combination is intended by the composer. In contrast, one would not normally try to make sense of, for instance, two random sounds on the street, since those sounds are usually not meant to be interpreted together. Though relatively peripheral to the emergence of composite meaning, the notion of the intentionality of alignment is indispensable in the analysis of multimodal meaning construction, especially in cases where incongruity is involved, as the assumption about the intentionality of alignment—or lack thereof—

determines whether there will be emergent meaning in the first place, and if positive, when and where to look for vital relations in between inputs.

Having identified some of the notions vital for capturing the complex, dynamic nature of multimodal interactions where the relation of expectation is involved, in the following I further demonstrate how these theoretical notions can be put together to capture the dynamic meaningfulness directly or indirectly resulting from expectation.

## 3.2. ASSEMBLING THE TOOLS.

Given the complex nature of multimodal meaning construction, any legitimate attempt to account for relevant phenomena is necessarily convoluted and is not easily reducible to some overly idealized, discrete parameters, despite the various commonalities shared by the different modes. Incorporating the key theoretical notions just reviewed, in the following I will demonstrate how the complexity of multimodal meaning construction can be captured comprehensively and in a coherent way. The toolbox captures in a single unified framework meanings expressed through different modes, the progression of time, conceptual distances between mental spaces, and processes of conceptual integration and disintegration, among others. In view of the complexity of the toolbox and the number of dimensions involved, a visualized representation is presented in Figure 2 to provide a clearer illustration.

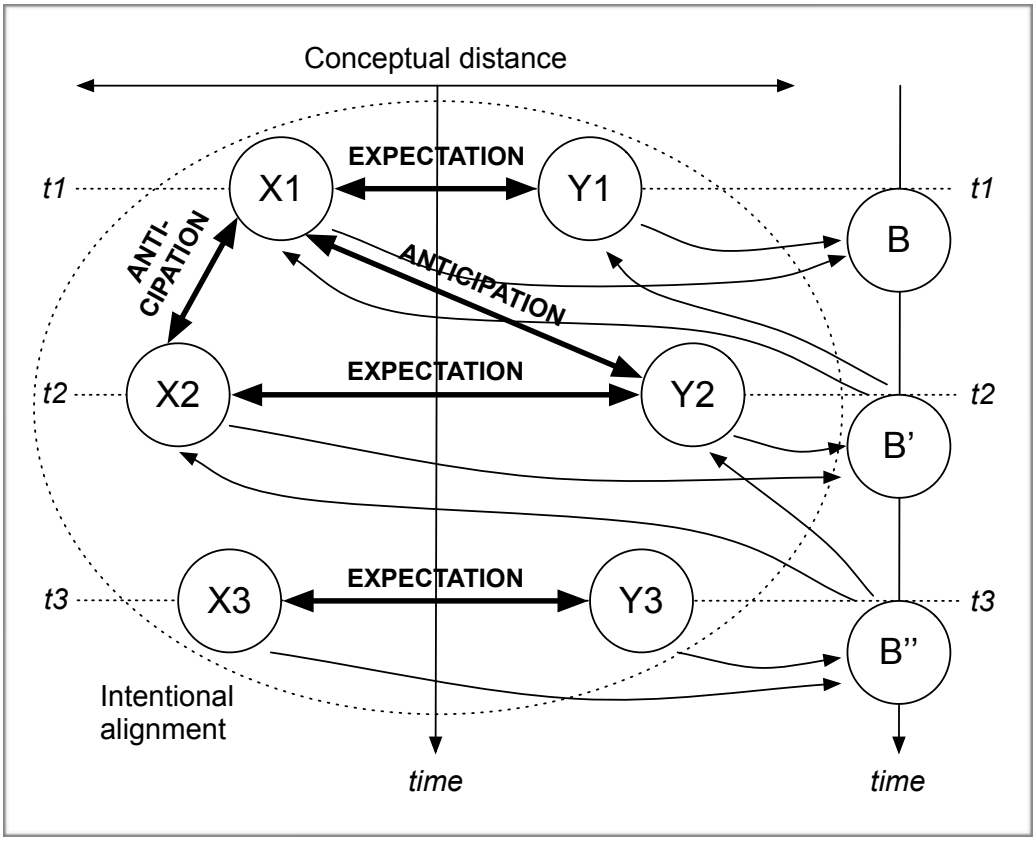


Fig. 2. Toolbox for dynamic multimodal meaning construction.

Starting from the axes, the horizontal one in Figure 2 represents the conceptual distance between constructions occurring at the same time. It is worth noting again that the idea of conceptual distance is a relative one, as there is no way to determine the exact and absolute conceptual distance between different meanings. The vertical axis, on the other hand, represents time. Although two axes of time are shown in Figure 2, there is really only one. The reason the axis representing time is presented twice is because the toolbox is really multidimensional (in the case of Figure 2, three-dimensional), and that, in order for it to fit in a two-dimensional representation, some

dimensions are collapsed for a clearer visual presentation, resulting in the repetition of the axis of time. Along these axes are constructions from modes X and Y and their convoluted interactions with one another.

At t1, a point in time, there exists a construction in mode X—X1—and a construction in mode Y—Y1. Between them is Expectation, the vital relation that allows the listener to realize that what they are listening to is not just some random combination of language and music but a single unified piece. Elements and relations in X1 and Y1 are selectively projected into a blend, B, which takes shape slightly after t1. Through composition, completion, and elaboration—the three key steps of meaning emergence—blended meaning emerges.

At t2, there are also two constructions, X2 and Y2, except that the conceptual distance in between has significantly increased. The longer distance between the two constructions imposes tension on Expectation, the intermodal relation that holds between them. Elements and relations in X2 and Y2 are selectively projected into B, transforming it into B'. While different constructions in modes X and Y exist at different points in time, B does not get replaced over time. Instead, it only keeps getting elaborated and modified. As new meaning has emerged in B', some of the emergent meaning gets disintegrated back into X1 and Y1—or more precisely, the listener's memories of X1 and Y1—modifying the listener's initial understanding of them. The same process is repeated at t3, and then t4, t5, and so forth. Besides

Expectation, which holds between constructions represented simultaneously, there is also the vital relation of Anticipation between constructions that are sequentially related, whether inter- or intra-modally; for instance, between X1 and X2, and X1 and Y2. Finally, all these constructions are within the boundaries of intentional alignment; that is, the temporal alignment of these multimodal constructions is intended, so that the cognizer is prompted to make sense of the alignment.

It ought to be noted that the toolbox is presented in Figure 2 in an intentionally clear and neat way that is easier for the reader to comprehend, and that the actual framework is much more convoluted. To start with, the intervals between different points in time do not have to be even and can vary greatly. The cognitive operation of conceptual disintegration is also much more variable than the figure is able to capture. For instance, meaning emergent in B'' may be disintegrated and projected back not only into X2 and Y2 but also into X1 and Y1, and X3 and Y3, modifying the cognizer's initial understanding of the constructions preceding B''. The fact that only two modes are present in Figure 2 is likewise a simplification. Cases of multimodal communication involving more than two modes can be easily spotted, as is often the case in, say, movies, where linguistic, musical, and visual constructions often take place simultaneously. The multidimensional nature of the toolbox allows the number of dimensions in the framework to adapt accordingly.

In accordance with the tenets of construction grammar, the toolbox imposes little limitation on the formal scope of constructions. For instance, while at t1 the cognizer comprehends and compares X1 and Y1, at t2 they do the same not only for X2 and Y2 but also for X1+2 (the larger unit made up of X2 and the listener's memories of X1) and Y1+2 (the larger unit made up of Y2 and the listener's memories of Y1). Likewise, at t3, the cognizer repeats the same process for X1+2+3 and Y1+2+3, X2+3 and Y2+3, and of course, X3 and Y3. *Eo ipso*, X1, for instance, can itself be a composite structure of smaller, component constructions. Essentially, it is in most cases the gestalt, holistic meaning that the cognizer can and does comprehend, compare, and make sense of.

In sum, the point of the visual presentation in Figure 2 is to show what dimensions need to be considered, what elements and relations are at work, what directions the relations go, and so forth. Despite the static representation, the whole system captured in the toolbox is dynamic. The balance between the numerous relations in the toolbox oscillates from one point in time to the next, contributing to the dynamic meaningfulness of the entire passage.

Having demonstrated what can be captured in the toolbox and what it is capable of, in the following chapter I will revisit the aforementioned *Siegfried* passage, to show how the unexpectedness in the passage can be captured and analyzed using the toolbox. While the toolbox is perfectly capable of accounting for modalities other than

language and music, due to the scope of the present study, only cases of language-music combination will be examined. Given that the present study aims mainly at proposing a prospective theoretical framework within which to investigate multimodal meaning construction, the following analysis will also only offer a sketch of how the toolbox can be applied to real-life examples. It is hoped that the analysis in the present study can offer insight for further investigation and serve as a reference on which future empirical research, including studies on other sorts of multimodal means of communication, can draw.

## 4. The dishonest Mime and his language-music mismatch.

*Siegfried*, the second day (third part) of Richard Wagner's epic tetralogy *der Ring des Nibelungen*, tells the story of Siegfried, the son of Siegmund and Sieglinde, growing into an adventurous young adult who does not know the meaning of fear. An orphan of the Wälsung race, Siegfried is brought up in the woods by Mime the Nibelung, who is aware of Siegfried's potential to obtain the Ring from Fafner the dragon, and who has raised baby Siegfried with reluctance only so that adult Siegfried would obtain the Ring for him. The duality of love and hatred in the deceitful nature of Mime's attitude towards Siegfried is capitalized on by Wagner, who through extensive use of leitmotifs showcases the effects brought about by the dynamic intermodal relation between simultaneously represented language and music.

A leitmotif can be defined as a short musical figure that is associated with a character, an object, an event, or an emotion (Sabor 1997a). While not the first one to make use of leitmotifs in musical composition, Wagner is certainly the first and likely the only one to use leitmotifs so extensively and in such an intricate manner, deploying leitmotifs as a key component of his ideal of *Gesamtkunstwerk*—the totality of the arts—which seeks to fuse media such as music, language, drama, acting, etc. into a single, unified artistic experience. Wagner's leitmotifs serve a plethora of functions that are far beyond the purely musical; as Sabor (1997a: 137) aptly puts it, "they underpin the action, they comment on it, they help to create receptive moods in the listener, they

elucidate, they sometimes tell the audience what the characters on stage do not yet know, they prophesy, and they occasionally contradict the evidence before our eyes." In this sense, Wagnerian leitmotifs fit perfectly into the definition of constructions, as each leitmotif is a pairing of form (in most cases a short tune and/or a sequence of chords) and function (the association with a character, object, event, or emotion).

The omnipotence of Wagner's leitmotifs makes it possible for stories to be told by just leitmotifs. In fact, one can even go so far as to say that the music of his mature music dramas, *Siegfried* included, is nothing but an extremely intricate tapestry of intimately interwoven leitmotifs. As Wagner writes himself to a friend, "There is hardly a single bar in the orchestra [of *der Ring des Nibelungen*] which is not based on preceding motifs" (cited in Sabor 1997a: 147). Probably also because of the omnipotence of his leitmotifs, Wagner believes in the self-explanatory power of leitmotifs, and does not explicate the meaning of most of his leitmotifs. Indeed, the attentive listener would usually have no problem associating leitmotifs with their corresponding concepts, objects, events, etc. Regardless, numerous music critics and musicologists have attempted to provide systematic accounts of the leitmotifs, notably Donington (1976), Herbert (2008), and Sabor (1997b).

The following are two of *Siegfried*'s notable passages in which the rich potential of language-music incongruity is remarkably exploited by the composer to produce effects in the listener. The analysis will start with an examination of the text and the

music of the passages. While the text will be broken down into episodes, the music will be analyzed in terms of leitmotifs, which I take to be the major musical constructions in the Wagnerian passages. Episodes and leitmotifs are certainly not the only constructions in the passages—for instance, meaning may exist under or above the episode level, and there may be noteworthy progressions of harmony on top of the leitmotifs—but the following analysis will be focused on the linguistic episodes and musical leitmotifs, as they are the levels on which most of the cross-modal interactions occur. Treating leitmotifs as the basic-level (Rosch and Lloyd 1978) constructions is also justified by the fact that Wagnerian leitmotifs emerge through usage. That is, the association between the form and function of a leitmotif is not stipulated by the composer, but has to be made by the listener in the course of listening to the music drama.

Since available accounts of leitmotifs identify different leitmotifs and use different labeling schemes, for consistency the present study will draw primarily on Sabor's (1997b) identification and categorization, unless otherwise specified. In cases where the leitmotif in question is not identified by Sabor (1997b), Donington's (1976) and Herbert's (2008) labeling schemes will be referenced, in that order. Importantly, in examining Wagner's work, one should always bear in mind that what is important about the leitmotifs is their relation to their associated characters, objects, events, or emotions, instead of the linguistic labels assigned to them, as those labels are created

and attributed to them by music critics and musicologists mainly to assist in the listener's understanding or to facilitate analysis.

Having had their text and leitmotifs examined, each of the passages will then be analyzed within the framework of the toolbox. Since the focus of the present study is the dynamicity of meaning construction in the passages, analyses of synchronic sections, such as detailed accounts of synchronic conceptual integration networks of particular moments[6], will not be provided.

A final issue pertaining to the following analysis of Wagner's work is that attention will only be paid to events taking place in the imaginary world of *der Ring des Nibelungen*. Due to limitations of scope and the focus of the present study, meaning related to the social and political context of the time of composition, though key to a well-rounded understanding of the work, will not be addressed.

## 4.1. MIME'S COMPLAINT REVISITED.

The First Scene of the First Act of Wagner's music drama *Siegfried* opens with a conversation between the young hero Siegfried and his "foster father" Mime. Seeing Mime's failed attempt at forging a sword, the pompous Siegfried mocks the old dwarf condescendingly. Baffled by Siegfried's disrespect and ungratefulness, Mime sings the

[6] For such analyses of other works, see for instance Hsu and Su (2014) and Zbikowski (1999, 2002).

famous "Als zullendes Kind zog ich dich auf," the passage briefly discussed in Chapter 1.

### **4.1.1.** "ALS ZULLENDES KIND ZOG ICH DICH AUF."

The following are the text and the English translation (Spencer and Millington 1993) of "Als zullendes Kind zog ich dich auf," starting four lines before the exact passage. The text can be seen as consisting of three episodes: Starting with some rather sarcastic remarks about Siegfried's ungrateful and arrogant attitude in the first episode (lines 1–4), in the second episode (lines 5–32) Mime's words turn into a monologue in which he stresses how much effort he has devoted to raising Siegfried, how much sacrifice he has made for Siegfried, how much care and love he has always had for Siegfried, and the like. Towards the end, in the final episode (lines 33–36), Mime's words once again turn slightly ironic, as he complains about Siegfried's lack of appreciation of his devotion.

**MIME**
Das ist nun der Liebe
schlimmer Lohn!
Das der Sorgen
schmählicher Sold! —

Als zullendes Kind
zog ich dich auf,
wärmte mit Kleiden
den kleinen Wurm:
Speise und Trank
trug ich dir zu,
hütete dich
wie die eig'ne Haut.
Und wie du erwuchsest,
wartet' ich dein;
dein Lager schuf ich,
dass leicht du schlief'st.
Dir schmiedet' ich Tand
und ein tönend Horn;
dich zu erfreu'n
müht' ich mich froh:
mit klugem Rathe
rieth ich dir klug,
mit lichtem Wissen
lehrt' ich dich Witz.
Sitz' ich daheim
in Fleiß und Schweiß,
nach Herzenslust
schweif'st du umher:
Für dich nur in Plage,
in Pein nur für dich
verzehr' ich mich alter
armer Zwerg!

Und aller Lasten
ist das nun mein Lohn,
daß der hastige Knabe
mich quält und haßt!

**MIME**
That's the sorry
wages of love!
That's the shameful
reward for my cares!

From a suckling babe
I brought you up,
warmed the little
mite with clothes:
food and drink
I brought to you
and tended you
like a second self.
And when you grew bigger
I waited upon you;
I made you a bed
so you'd sleep more softly.
I forged for you toys
and a winding horn;
to give you pleasure
I gladly toiled:
with clever counsel
I counselled you cleverly,
with lucid lore
I taught you wit.
While, toiling and sweating,
I sit at home,
you roam around
to your heart's content:
suffering torment for you alone,
for you alone I suffer affliction
and wear myself out, a poor
old dwarf!

And that's my reward
for the burdens I've borne,
that the quick-tempered boy
torments and abhors me!

As for the music accompanying these words, the leitmotif identified by Donington (1976) as "Mime as Obsessive Persistence," shown in Figure 3, is heard in the orchestral accompaniment throughout the passage.

Fig. 3. Leitmotif "Mime as Obsessive Persistence" (Donington 1974).

This leitmotif is derived from the "Forge" leitmotif identified by Sabor (1997b). It is a motif tightly associated with Mime's character that is first heard at the start of *Siegfried*, when Mime attempts to forge a sword:

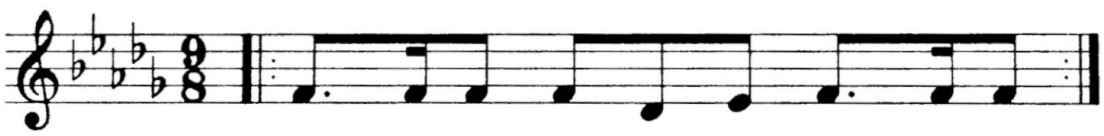

Fig. 4. Leitmotif "Forge" (Sabor 1997b).

"Mime's Complaint[7]," a leitmotif closely related to the "Mime as Obsessive Persistence" motif, is identified and labeled by Donington (1976) in Mime's sung melody accompanying his second episode of text, as shown in Figure 5.

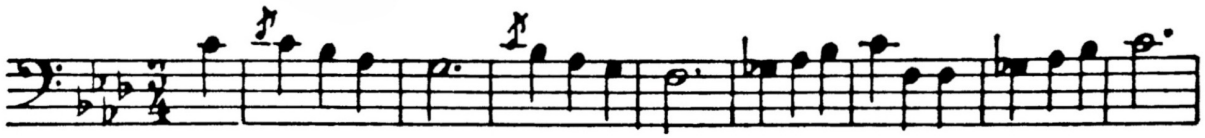

Fig. 5. Leitmotif "Mime's Complaint" (Donington 1974).

Besides the leitmotifs of "Mime's Complaint" and "Mime as Obsessive Persistence," which are heard almost throughout the entire passage, Sabor (1997b) also identifies a motif heard accompanying the third episode of text, at lines 33–34 ("Und aller Lasten ist das nun mein Lohn"), which he calls the "Grief" leitmotif:

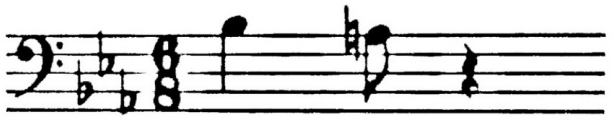

Fig. 6. Leitmotif "Grief" (Sabor 1997b).

[7] The first five bars of the "Mime's Complaint" leitmotif as Donington (1976) calls it are also identified as a motif by Sabor (1997b), who labels it "Crocodile." Since the name "Crocodile" is very similar to "Crocodile Tears," a label Herbert (2008) assigns to a key leitmotif in the passage to be discussed in Section 4.2, here Donington's (1976) label "Mime's Complaint," instead of Sabor's (1997b) "Crocodile," is used to avoid confusion.

In short, whereas the text of "Als zullendes Kind zog ich dich auf" can be broken into three fairly clear-cut episodes, the accompanying music is mostly filled with the "Mime's Complaint" and the "Mime as Obsessive Persistence" leitmotifs, except towards the end of the passage, where the "Grief" motif is heard.

### **4.1.2.** Pretense unveiled.

Analyzed as a whole using the toolbox from Chapter 3, the passage of simultaneously represented language and music can also be roughly cut into three episodes: the episode before "Als zullendes Kind," the episode starting from "Als zullendes Kind," and the one starting from "Und aller Lasten." For the first episode, relation between the meaning of the text and that of the music is rather neutral: While the text expresses Mime's negative emotions about Siegfried's lack of appreciation for all the hard work Mime has done for him, the music is mainly a leitmotif closely associated with Mime's character. This suggests a short intermodal conceptual distance. As time progresses and the passage proceeds to the second episode, where "Als zullendes Kind zog ich dich auf" officially starts, the semantics of Mime's words turn into a monologue where Mime goes into detail about how selfless and loving he has been in taking care of Siegfried, while the leitmotif of "Mime's Complaint" still sounds in Mime's tune. The mismatch between the information from the two modes signals a greater intermodal conceptual distance, which in turn imposes great tension

on Expectation, the vital relation that holds the two modes together as a single unified piece of work. As Mime's monologue reaches its end, the listener is presented with the third episode of language-music combination. Here, the language once again turns into a grudging complaint about Siegfried's ungratefulness and egoism, echoing the accompanying leitmotif of "Grief," which further reinforces the negativity of the text. This leads the intermodal conceptual distance between the language and the music to shrink, easing the tension on the vital relation of Expectation.

Besides the dimensions of time and conceptual distance, the dimension on which the input spaces interact with the blend also undergoes dynamic changes throughout the passage. Of the three episodes of the passage, the second one, where there is mismatch and where the intermodal conceptual space is on average the longest, sees some particularly intriguing dynamics. For instance, the listener might not be able to make sense of Mime's true meaning when he first sings the phrase "Als zullendes Kind zog ich dich auf" to his complaining leitmotif. As time progresses, the listener may at a later stage get the sense of Mime being pretentious from new elements and relations emerging from the ever-evolving blended space. Some of these emergent structures then get projected back into earlier linguistic and musical input spaces, elaborating and modifying, for example, the meaning of the "Als zullendes Kind zog ich dich auf" phrase, which may not have been properly interpreted by the listener at first. These intertwining cognitive processes are at work the whole time,

contributing to the dynamic meaningfulness of the entire passage, and ultimately to the entire music drama.

In the following, I will turn to a later passage in the same music drama. While there are also dynamic changes in cross-modal conceptual distance, in this case the roles of language and music are "reversed."

## 4.2. Crocodile tears.

After the quarrel in the passage examined above, Siegfried successfully forges together the fragments of Nothung, an enchanted sword once owned by his parents Siegmund and Sieglinde. With Nothung now reforged into one piece, Siegfried slays Fafner the dragon with ease, obtaining the Ring guarded by the dragon. As Siegfried licks his fingers, on which some of the dragon's blood has been spilled, he realizes that he can now understand the language of nature, which takes the form of the Woodbird, and that he is able to decipher the true intention underlying other people's words. In the meantime, now that Siegfried has obtained the Ring from Fafner the dragon, Mime's master plan, which has been to raise Siegfried and obtain the Ring through Siegfried's strength, is also reaching completion. The final step yet to be taken is obtaining the Ring from Siegfried, and Mime chooses to do so by poisoning Siegfried with the drugged potion he has prepared in the First Act of the music drama. To trick Siegfried into drinking the potion, Mime again tries to lure him with caring words and

sweet tunes. However, thanks to the supernatural power Siegfried has newly acquired from tasting the dragon blood, while the music is still in the affectionate and sweet tune Mime intends him to perceive, to Siegfried's (and also to the listener's) ears, Mime's words are "not the words actually spoken by the dwarf but the unspoken thought at the back of his mind" (Newman 1949: 609), which is to take Siegfried's life. Unaware of the fact that the deceptive words he has carefully crafted are being "translated" into his true intention for Siegfried, Mime reveals unknowingly all the details of his evil masterplan. In revulsion, Siegfried slays Mime with a quick blow of Nothung (Herbert 2002, 2008; Millington 2006).

### **4.2.1.** "WILLKOMMEN, SIEGFRIED."

Excerpted from Act 2, Scene 3 of *Siegfried*, the passage "Willkommen, Siegfried" is one of Wagner's "masterpieces of characterization" (Newman 1949: 607), where Mime attempts to lure Siegfried into consuming the poisonous potion, but where he sings his murderous intention to "tones of deepest flattery and affection" (Donington 1974: 197). Indeed, as Wagner explains himself, in this passage Mime "blurts out the truth in words which contrast with the honeyed music—a device which is very comical" (cited in Sabor 1997b: 141). At least at first glance, this is precisely the opposite of what happens in "Als zullendes Kind zog ich dich auf," in which it is the music instead of the language that gives away Mime's evil intention (but see the

discussion in Chapter 5). The German text and the English translation (Spencer and Millington 1993) are given in the Appendix. Though consisting of numerous alternating turns between Mime and Siegfried, the passage consists essentially of repetitions of a mostly fixed pattern. The following excerpt, taken from near the start of the passage, clearly exemplifies this pattern:

| | | |
|---|---|---|
| | **MIME** | **MIME** |
| | Nur sachte! Nicht lange | But soft! You'll not |
| 15 | sieh'st du mich mehr: | have to see me much longer: |
| | zum ew'gen Schlaf | I'll soon lock |
| | schließ' ich dir die Augen bald! | your eyes in lasting sleep! |
| | Wozu ich dich brauchte, | You've done |
| | hast du vollbracht; | what I needed you for; |
| 20 | jetzt will ich nur noch | all that I still want to do |
| | die Beute dir abgewinnen: — | is to win from you the booty: — |
| | mich dünkt, das soll mir gelingen; | I think that I ought to succeed in that; |
| | zu bethören bist du ja leicht! | you're easy enough to fool after all! |
| | | |
| | **SIEGFRIED** | **SIEGFRIED** |
| | So sinn'st du auf meinen Schaden? | So you're planning to do me harm? |
| | | |
| | **MIME** | **MIME** |
| 25 | Wie sagt' ich denn das? | What, did I say that? |
| | Siegfried, hör' doch, mein Söhnchen! | Siegfried, sonny, listen to me! |
| | | |
| | Dich und deine Art | You and your kind |
| | haßt' ich immer von Herzen; | I have always hated with all my heart; |
| | aus Liebe erzog ich | it was not out of love |
| 30 | dich lästigen nicht: | that I brought you up, you burdensome child: |
| | dem Horte in Fafner's Hut, | my efforts were aimed at the gold, |
| | dem Golde galt meine Müh'. | at the hoard in Fafner's safekeeping. |
| | Gibst du mir das | If you don't give it up |
| | gutwillig nun nicht: — | to me willingly now, |
| 35 | Siegfried, mein Sohn, | Siegfried, my son, |
| | das sieh'st du wohl selbst — | you must see for yourself — |
| | dein Leben mußt du mir lassen! | you must yield up your very life to me! |

As the excerpt shows, Mime first declares that he wants to take Siegfried's life (lines 14–23). Hearing this, Siegfried asks Mime to confirm what he has just said (line 24). Mime then denies his having said he wants to kill Siegfried (lines 25–26), but then goes on reiterating his intention to take Siegfried's life (lines 27–37). In other words,

the pattern consists of three episodes: Mime saying he wants to kill Siegfried, Siegfried confronting Mime with his "slip of tongue," and Mime rejecting Siegfried's accusation.

While the text of most parts of "Willkommen, Siegfried" can be seen as repeated cycles of these three episodes, one noteworthy exception is found at lines 43–51 (see Appendix). The music accompanying these words is the "Mime's Complaint" motif. This part will not be discussed in the present section, as its language-music relation highly resembles that in the second episode of "Als zullendes Kind zog ich dich auf" (cf. Section 4.1).

With regard to music, although numerous leitmotifs are heard accompanying the conversation between Mime and Siegfried, two leitmotifs dominate the passage, the first of which, shown in Figure 7, is identified and labeled the "Crocodile Tears" motif by Herbert (2008):

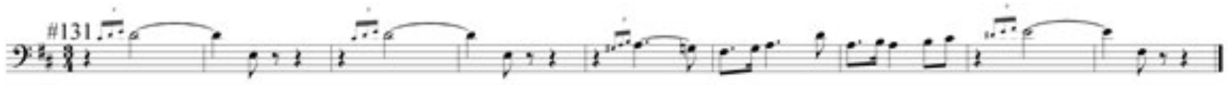


Fig. 7. Leitmotif "Crocodile Tears" (Herbert 2008).

Towards the end of the text excerpt above, starting from line 35 ("Siegfried, mein Sohn"), a transformed, more dramatic version of the leitmotif is heard, as noted by Newman (1949):

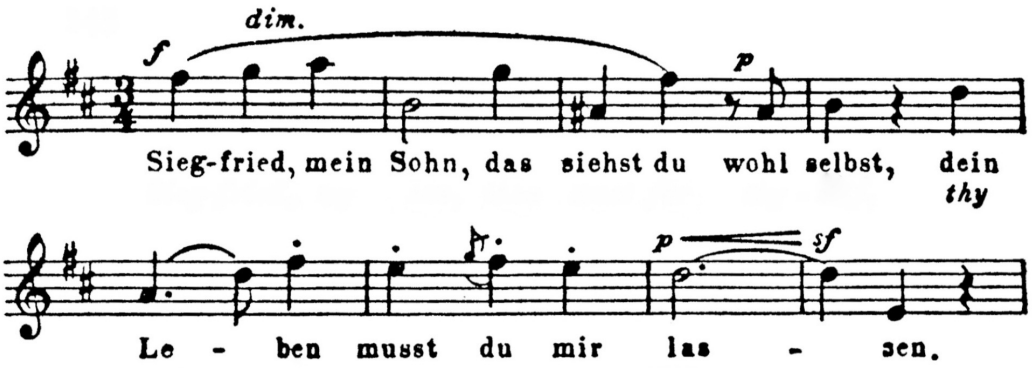


Fig. 8. Excerpt 143 from Newman (1949).

Also labeled “False Flattery” (Herbert 2008), the “Crocodile Tears” leitmotif is the flattering and affectionate (Donington 1974) music into which Mime puts “all the honey he can” (Newman 1949: 607). While the labels of “Crocodile Tears” and “False Flattery” suggest that the meaning of the motif should not be taken at face value, the fact that one should not take the meaning of the motif at face value in order to understand the emergent meaning of the passage does not change the “literal” meaning of the motif. One should bear in mind that these labels are coined *post hoc* by people who have grasped the emergent meaning of the music drama. That is, the leitmotif is labelled this way because the people naming the leitmotif already understand that the flattery and affection in this motif are faked. The fact that the disintegrated emergent meaning makes the sweetness of the motif dishonest does not change the fact that the motif is extremely honeyed.

Apart from the “Crocodile Tears” motif, the other leitmotif dominating the passage is “Woodbird,” identified by Sabor (1997b):

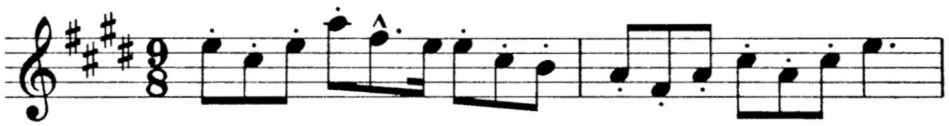

Fig. 9. Leitmotif "Woodbird" (Sabor 1997b).

This leitmotif is associated with the information the Woodbird has previously provided Siegfried, namely that, having tasted the dragon blood, Siegfried is now able to decipher the true, hidden intention underlying people's words, and that he should be wary of Mime's treacherous, evil scheme.

Putting language and music together, the "Crocodile Tears" motif is heard during the episode in which Mime expresses his intention to take Siegfried's life, as well as the episode in which Mime dismisses Siegfried's accusation, whereas the "Woodbird" leitmotif is heard where Siegfried questions Mime's intention.

The rotation of the three episodes continues for several cycles. With each rotation, Mime's words become more explicit, his music tenderer and oilier (Newman 1949). The semantic gap between Mime's language and music becomes extremely wide towards the end of "Willkommen, Siegfried," right before Siegfried decides he can no longer put up with Mime's slyness and slays him with Nothung:

| | | |
|---|---|---|
| | **MIME** | **MIME** |
| | Was möcht' ich? Sagt' ich denn das? — | I mean to do what? Is that what I said? — |
| | | |
| | Ich will dem Kind | I want only |
| | nur den Kopf abhau'n. | to hack the child's head off. |
| 100 | Denn haßte ich dich | For even if |
| | auch nicht so sehr, | I hated you less |
| | und hätt' ich des Schimpf's | and hadn't so much |
| | und der schändlichen Mühe | of your hateful abuse |
| | auch nicht so viel zu rächen: | and such shameful toil to avenge, |
| 105 | aus dem Wege dich zu räumen | I'd still waste no time |
| | darf ich doch nicht rasten, | in clearing you out of the way |
| | wie käm' ich sonst anders zur Beute, | for how else could I gain the spoils, |
| | da Alberich auch nach ihr lugt? — — | since Alberich covets them, too? — — |

At lines 98–99 ("Ich will dem Kind nur den Kopf abhau'n"), music so exaggeratedly sweet and passionate to the point that is almost comical is heard, as pointed out by Newman (1949):

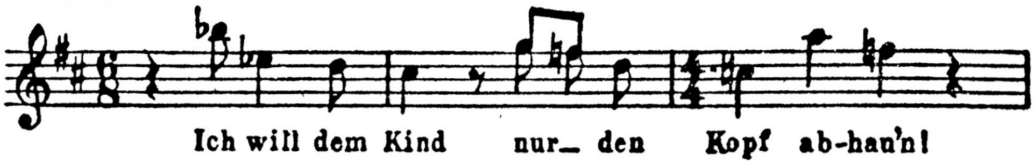


Fig. 10. Excerpt 144 from Newman (1949).

Starting from line 100 ("Denn haßte ich dich"), the music tones down from the highly dramatic sentiment. The listener now hears "the most honeyed tones" (Newman 1949: 609) Mime can assume—the last four notes of what Sabor (1997b) identifies as the "Longing" leitmotif, which has previously appeared where

Siegfried was recounting the beautiful love he had observed among animals in the forest, expressing his longing for a pair of caring parents.

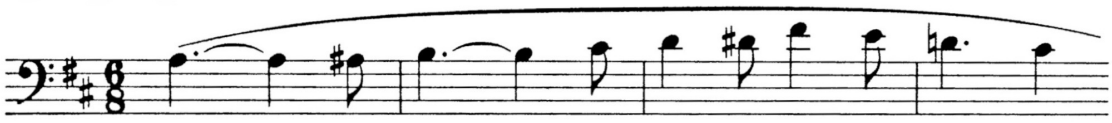

Fig. 11. Leitmotif "Longing" (Sabor 1997b).

The tension between Mime's contradicting language and music reaches a peak here. Having fully understood Mime's conspiracy, Siegfried kills the dwarf with a determined blow of the sword Nothung.

### **4.2.2.** ROLES REVERSED.

Looking at the language and music at the same time, the continual change in conceptual distance between the two modalities looms large. Of the three recurring episodes, the language and the music are conceptually closest to each other where Siegfried questions Mime's true intention and where the "Woodbird" leitmotif dominates the music, as it is the Woodbird's warning that helped Siegfried see through Mime's pretense in the first place. The intermodal conceptual distance is stretched where Mime denies Siegfried's suspicion and where the "Crocodile Tears" motif enters. It should be noted that, although it is clear to the listener that the "Crocodile Tears" motif aptly captures Mime's pretense, the honeyed tune of the leitmotif is actually

incongruous with the literal text, in which Mime defends his innocence and is well-meaning. The understanding that there is pretense in Mime's words is the result, rather than the cause, of the listener's cognitive processing. The parts where Mime sings his evil intention to the cheerful, happy tune is where the language and the music are conceptually furthest away from each other, as his words and the accompanying music directly contradict each other.

The length of the intermodal conceptual distance oscillates as the passage unfolds and the three episodes rotate. Where there is mismatch between language and music, such as where the sweet music accompanies Mime's description of how he will end Siegfried's life, the longer intermodal conceptual distance imposes tension on the vital relation of Expectation. This in turn prompts the listener to make sense of the seemingly anomalous multimodal construction through conceptual blending. With the contextual knowledge that Mime has always wanted to obtain the Ring through Siegfried's strength and that Siegfried is now capable of understanding people's true intentions even beneath their deceptive words, the listener is able to come up with the emergent meaning that Mime is really trying to poison Siegfried in order to get the Ring from him. Likewise, though not as incongruous, the simultaneous representation of the "Crocodile Tears" leitmotif and Mime's defending himself also stretches the intermodal conceptual distance, prompting the listener to come up with the emergent understanding that the "Crocodile Tears" motif is used to mark the dishonesty of

Mime's words, even though the leitmotif itself does not explicitly give any information about Mime's true intention underlying his dishonesty. Conceptual blending is at work even when incongruity is minimal or non-existent. For instance, to make sense of the temporal juxtaposition of the "Woodbird" leitmotif, which signals Siegfried's ability to see through Mime's deceptive words, and Siegfried's questioning of Mime, the listener needs to draw upon their contextual knowledge—more specifically their knowledge of the story of the music drama prior to the present passage, that Siegfried is able to understand the Woodbird's warning against Mime after tasting the dragon blood—to come to the emergent understanding that the composer intends to remind the listener of Siegfried's ability to decipher people's real intentions, and that it is because of this ability that Siegfried confronts Mime.

The intermodal dynamics of "Willkommen, Siegfried" differ from that of "Als zullendes Kind zog ich dich auf" in many ways, the most striking being that, although in "Als zullendes Kind zog ich dich auf" it is the music that gives away the information Mime tries to mask with his words, it is not the case in "Willkommen, Siegfried." Where Mime sings his intention to end Siegfried's life to the sweet tune, the tune communicates what he intends Siegfried to believe, and it is his language that betrays him. Where Mime tries to defend himself after Siegfried's confrontation, neither the language nor the music is "telling the truth," since the accompanying "Crocodile Tears" leitmotif provides no information about the truth, merely suggesting Mime's

dishonesty. In other words, the "roles" of language and music are reversed in "Willkommen, Siegfried": Where there is intermodal incongruity, it is the language instead of the music that betrays the character and "reveals the truth"—if any one of the modalities is revealing the truth, that is.

The complexity of "Willkommen, Siegfried" is however much more than the "reversed roles," which is precisely why it is more difficult and counter-intuitive to determine which mode is "revealing the truth." The following chapter therefore attempts to disentangle this complexity by delving deeper into the intermodal relations in the two *Siegfried* passages.

## 5. THE MANY COATINGS OF MISMATCH.

The intricacy of the cases of language-music mismatch in the two *Siegfried* passages makes them irreducible to mere instantiations of irony, sarcasm, dishonesty, or anything else. This is especially true for the second passage, in which the cunning Mime actually ends up "revealing the truth," but in a way that does not feel like he is really telling the truth. Such complexity calls for deeper contemplation of the inter-modal relation between the language and the music in the two passages, as well as the social-interactional dynamics between the two agents of Siegfried and Mime. To understand what makes the phenomena in these passages extraordinary, one needs to first identify the norms from which these phenomena deviate. In the previous sections, I have argued that the reason why language-music incongruity stands out is that it violates the listener's default expectation that constructions simultaneously represented should carry similar or complementary semantics. I have also identified the theoretical notions necessary for the analysis of multimodal mismatch and put forth a toolbox that has provided insights about the two *Siegfried* passages. Nonetheless, it appears that the theorizing and analysis so far, which are primarily cognition-oriented, are not sufficient to capture all the nuances of the passages, not least those in "Willkommen, Siegfried."

As Croft (2009, 2011) points out, cognitive linguistics as a general theory of linguistics needs to be able to incorporate and account for the social-interactional

aspect of linguistics, which is a necessary, integral part of language use, and which is one of the primary reasons why humans need language in the first place. Indeed, with the amount of emphasis cognitive linguists lay on real-life language use—a reaction to generative approaches that examine linguistic phenomena in a theoretical vacuum free from the "noise" of real usage events—it is hard, if not impossible, for cognitive linguists to make a convincing case without being able to at least address the social-interactional aspect of language. Though currently limited compared to research in other strands of cognitive linguistics, several attempts in this direction has been put forward, such as Chafe (1997a, b), Clark (1996), and Du Bois (2007, 2014).

Rooted in relevant philosophical investigation on joint intention (e.g. Bratman 1987, 1997; Lewis 1969), Clark's work tackles some of the most fundamental issues on the social-interactional side of language use. In particular, he argues that "people use language for doing things with each other" (Clark 1996: 387). That is, people use language to coordinate with each other in order to carry out what Clark calls joint actions, which include for instance making a transaction, planning an event, playing a game, chatting, and also using language itself. Clark argues further that it is possible for people to perform more than one joint action simultaneously, as a way to demonstrate events in a hypothetical world. When this happens, the joint actions taking place simultaneously can be seen as events occurring on different layers. For instance, in performing the two *Siegfried* passages, the tenor playing Siegfried's role and

the tenor playing Mime's part carry out two joint actions at the same time. On the first layer, which is situated in reality, they are two tenors singing to each other; on the second, fictional layer, they are Siegfried and Mime communicating with each other in the context of Wagner's music drama. In complex cases of language use, there can be even more joint actions simultaneously, and therefore even more layers. Importantly, for layering to function, participants' imaginations, as well as their mutual acknowledgment of the existence of the layers, is required. In Clark's view, it is because of language users' ability to carry out layering that some of the most intriguing phenomena in language, such as storytelling, irony, joking, etc., are possible.

Reexamining the two *Siegfried* passages in light of this, it appears obvious that no joint actions exist in the interaction between Siegfried and Mime. More precisely, each of them attempts at some points to conduct joint actions with the other person but fails: Siegfried is honest and straightforward in both passages, trying to establish common ground with Mime, whereas Mime constantly tries to trick Siegfried. While Mime's attempt may seem like a layer of pretense, it is actually not, since the existence of a layer requires that both parties acknowledge the layer. In the two passages, Siegfried either does not acknowledge the layers Mime attempts to construct, or simply is not aware of them at all. The only real layer of pretense in Clark's sense is the basis on which Siegfried and Mime's interaction in the two passages takes place. It is the layer of pretense jointly constructed by the people involved in the creation, the

production, and the appreciation of Siegfried and Mime's interaction—namely the audience, the composer and librettist Richard Wagner, the opera singer and actors, the orchestra, the conductor, the producer, the director, etc. of the music drama. These people jointly pretend that the events taking place onstage are real, that Siegfried and Mime are having a conversation in the context of the opera, and not merely that two singers are singing onstage to the accompaniment of the orchestra offstage.

Now the question remains what it is that happens between the roles of Siegfried and Mime, or more precisely, on the layer of the pretense jointly constructed by the audience and the performers, how the social-interactional dynamics between Siegfried and Mime that are not jointly constructed and appreciated can be made sense of to address the peculiarities of the two passages that the previous sections have not been able to account for.

In the case of "Als zullendes Kind zog ich dich auf," on the fictional Layer 2 jointly constructed by the audience and the performers, meanings associated with Mime are twofold. There is a meaning in his unspoken intention—that he does not care about the well-being of Siegfried—and a meaning in his spoken words—that he loves Siegfried and would sacrifice anything for him. These two meanings can be thought of metaphorically as two coatings of meaning, with Mime's intention on the first, more basic coating, and the meaning of his spoken words on the second coating. The second coating is on top of the first coating and covers it, so that under normal circumstances

only the second coating is accessible by people other than Mime himself. The process of transformation that "paints" the second coating on top of the first coating is done by Mime's evil scheme, in the sense that Mime "translates" his true intention into something else so as to convince Siegfried of his lie. These processes are summarized and visualized in Table 1:

<table>
<tr><td rowspan="3">Layer 2: Siegfried and Mime interact in the world of Wagner's der Ring des Nibelungen.</td><td>Coating 2 of meaning—Mime's spoken words: Mime genuinely cares about Siegfried, with no complaint.</td></tr>
<tr><td>↑<br>Mime's scheme<br>|</td></tr>
<tr><td>Coating 1 of meaning—Mime's intention: Mime has raised Siegfried grudgingly since he is only after the Ring.</td></tr>
<tr><td colspan="2">Layer 1: The listener, the singers, the orchestra, etc. jointly pretend that the events on Layer 2 are taking place.</td></tr>
</table>

Table 1. Meanings associated with Mime in "Als zullendes Kind zog ich dich auf."

From the listener's perspective[8], whereas Mime's words alternate between the first and second coatings of meaning, the music accompanying his words stays on the first coating. Recall that the passage can be roughly broken down into three episodes. In the first episode, Mime explicitly complains about Siegfried's ingratitude, before

[8] In opera, as is the case for most art or literal forms, because of the existence of the necessarily omniscient author and possibly an omniscient narrator, what the audience perceives is not always identical to what the characters perceive. This distinction will become even more vital in the discussion below on "Willkommen, Siegfried."

moving on to the second episode, in which he stresses how much love he has for Siegfried and enumerates the deeds he has done for Siegfried; in the third episode Mime recapitulates the first episode, becoming grumpy and sarcastic again. The whole linguistic passage can be seen as moving from the first to the second coating, before returning to the first coating. The music that is simultaneously represented, on the other hand, is dominated by such negative leitmotifs as "Mime as Obsessive Persistence," "Mime's Complaint," and "Grief," which are consistent with the semantics of the first coating.

The sense of mismatch arises in the second episode when Mime moves his speech up to the second coating but the music does not follow suit. Figuratively speaking, while Mime's words are already on the second coating, wherein we find the events he hopes Siegfried to believe in, his music stays on the first coating of meaning —the coating of his real intention. This explains not only the cross-modal mismatch but also why one can say that Mime's music "tips him off," revealing his true, malicious thought.

"Willkommen, Siegfried" can be analyzed likewise. On Layer 2, the pretense jointly put up by the audience and the performers, meanings associated with Mime are also twofold, including a meaning in his unspoken intention—that he intends to poison Siegfried in order to obtain the Ring—and a meaning in his spoken words (which is not accessible to the audience, or Siegfried)—that he intends to treat Siegfried to some

refreshment. These can also be seen as meanings on two different coatings, with a process of transformation "painted" by Mime on top of the first coating as part of his evil scheme. These are summarized and visualized in Table 2:

<table>
<tr><td rowspan="3">Layer 2: Siegfried and Mime interact in the world of Wagner's der Ring des Nibelungen.</td><td>Coating 2 of meaning—Mime's spoken words: Mime offers to treat Siegfried to a refreshment for fighting Fafner the dragon.</td></tr>
<tr><td>↑<br>Mime's scheme<br>|</td></tr>
<tr><td>Coating 1 of meaning—Mime's intention: Mime intends to kill Siegfried with the deadly potion.</td></tr>
<tr><td colspan="2">Layer 1: The listener, the singers, the orchestra, etc. jointly pretend that the events on Layer 2 are taking place.</td></tr>
</table>

Table 2. Meanings associated with Mime in "Willkommen, Siegfried."

From the listener's perspective, Mime's words again oscillate between the first and second coatings of meaning. The episode in which Mime pretends to be treating Siegfried to a refreshing drink is on Coating 2, while the episode in which he explicitly says he wants to poison Siegfried to death is apparently on Coating 1. The honeyed music that accompanies these words, on the other hand, mostly stays on Coating 2. The incongruity surfaces where the language and the music are on different coatings, namely where Mime sings his evil intention to the sweet tune.

While such an analysis of the "Willkommen, Siegfried" seems adequate at first glance, it does not spell out the aforementioned confusion of why it feels less intuitive

to determine which mode is "revealing the truth." To get to the bottom of this, one can start by spotting the crucial differences between the more ordinary "Als zullendes Kind zog ich dich auf" and the more convoluted "Willkommen, Siegfried." Although in both passages Mime's language alternates between voicing his true intention and reaffirming his faked love for Siegfried, it is only after "Als zullendes Kind zog ich dich auf" that Siegfried tastes the dragon blood, which in turn allows him to comprehend the language of nature, more specifically the speech of the Woodbird. That is, in "Willkommen, Siegfried," the supernatural power afforded by the dragon blood "translates" for Siegfried—and therefore for the listener—Mime's literal words into the true intention underlying those deceitful words.

The question then appears to be whether the "translated meaning" belongs on Coating 1 or a coating even higher than Coating 2. In other words, the problem is whether the dragon blood "peels off" Coating 2 and moves everything back onto Coating 1, or "paints" a new coating of meaning on top of Coating 2. Since what Siegfried—and therefore the listener—hears is the result of the dragon blood's "translation," which is not voiced by Mime himself and which Mime has no access to, it appears more reasonable to treat the "translated version" of Mime's words as occurring on a higher-level coating, one that is not identical to Coating 1, as shown in Table 3.

<table>
<tr><td rowspan="5">Layer 2: Siegfried and Mime interact in the world of Wagner's der Ring des Nibelungen.</td><td>Coating 3 of meaning—Mime's intention "translated": Mime intends to trick Siegfried into consuming the deadly potion in order to obtain the Ring.</td></tr>
<tr><td>↑<br>Supernatural power from the dragon blood and the Woodbird<br>|</td></tr>
<tr><td>Coating 2 of meaning—Mime's spoken words: Mime offers to treat Siegfried to a refreshment for fighting Fafner the dragon.</td></tr>
<tr><td>↑<br>Mime's scheme<br>|</td></tr>
<tr><td>Coating 1 of meaning—Mime's intention: Mime intends to kill Siegfried with the deadly potion.</td></tr>
<tr><td colspan="2">Layer 1: The listener, the singers, the orchestra, etc. jointly pretend that the events on Layer 2 are taking place.</td></tr>
</table>

Table 3. Alternative view on meanings associated with Mime in "Willkommen, Siegfried."

Re-examining the passage in this way, we can see that, from the listener's perspective, Mime's words really swing back and forth between the second and the third coatings of meaning, instead of between the first and the second. Unaffected by the effect of the dragon blood, the meaning of the accompanying music stays on Coating 2. The sense of incompatibility arises where Coating-3 language and Coating-2 music are simultaneously represented, in which case the cross-modal relation of expectation needs to hold together contradicting pieces of information.

Analyzing the "translated" version of Mime's words as an additional coating of meaning better illustrates the reason why it is more difficult and less intuitive to determine, in the case of "Willkommen, Siegfried," which of the two modes is

"revealing the truth": When Mime's words are on Coating 3, even though the meaning is very close to that on Coating 1, it is in fact the result of two processes of transformation, from Coating 1 to Coating 2 and then to Coating 3. This is unlike the music in "Als zullendes Kind zog ich dich auf," which, though also giving away Mime's true evil intention, operates directly on Coating 1. Adding to the complexity of "Willkommen, Siegfried" is that, while the language oscillates between Coatings 2 and 3, the meaning of the music stays on Coating 2. In other words, neither of the modes are directly grounded in Coating 1, which for the characters is the basic, literal coating of meaning that is not necessarily explicitly expressed or shared with other characters. This is also unlike the case of "Als zullendes Kind zog ich dich auf," in which the music is always anchored to Coating 1. The fact that, in "Willkommen, Siegfried," information about Mime's true intention comes not from the coating on which the true intention belongs but from two coatings above, together with the lack of any one of the modes anchored to the basic coating, makes the relationship among the linguistic meaning, the musical meaning, and Mime's true intention a lot more elusive than that in the more straightforward case of "Als zullendes Kind zog ich dich auf."

While the two passages differ in various ways, an interesting underlying commonality between them is that the music—whether it is a character's tune or the orchestral accompaniment—is in both cases only accessible to the audience. In "Als zullendes Kind zog ich dich auf," the listener can decipher Mime's duality and

dishonesty through the information provided by the accompanying music, while Siegfried, Mime's addressee, only hears Mime's spoken words, and has no access to Mime's evil intention. Similarly, in "Willkommen, Siegfried," neither character hears the crucial "Crocodile Tears" or "Woodbird" leitmotif, which the audience has easy access to. Siegfried is able to see through Mime's evil scheme only because what he actually hears is the "translation" done by the supernatural power afforded by the dragon blood.

In cases where multiple coatings of meaning coexist, it is crucial to distinguish the listener's perspective from, say, the character's perspective, because what the composer intends for the listener to perceive is not necessarily what the character also perceives. For instance, in "Willkommen, Siegfried," the composer intends for the listener to hear Mime's words through Siegfried—the Siegfried that is under the influence of the dragon blood—but what Mime hears himself is the words he "actually" utters, which the listener has no access to. Such distinction is especially crucial in analyzing cases where an omniscient author or narrator—one that can determine and alter the listener's perspective—exists, or where none of the channels of information is on the basic coating. It is precisely the composer's intricate manipulation of this distinction that makes the passage of "Willkommen, Siegfried" so unique and at the same time confusing for the analyzer.

Also crucial to the analysis of incongruity—multimodal or not—is the distinction between intention and effect. The main purpose of communication is to change other people's current knowledge structure, to create effects on other participants of a communication event (cf. Clark 1996; Grice 1957). The effect one wishes to have on other people is part of the person's intention. Put in another way, if one's intention is their goal, effect can be seen as their means to achieve that goal. However, despite being part of intention, effect is not always congruous with intention, at least not on the surface. For example, the intention of a liar is usually different from the effect they want on the addressee. The liar wants the addressee to believe something. The belief is the effect the liar wants to have on the addressee, but it is the fact that the addressee buys the lie, rather than the false belief itself, that is the liar's intention. Consider also the interaction between Siegfried and Mime in "Willkommen, Siegfried." Whereas Mime's intention is clearly to end Siegfried's life by poisoning him, the effect he wishes to impose on Siegfried is the belief that Mime is well-meaning and is offering him a refreshment as a reward for his hard work. In other words, even though the words Mime uses to trick Siegfried are at first glance incongruous with his evil intention, it is in fact compatible with the effect he intends to have on Siegfried, which in turn is part of his master plan to realize his ultimate intention. The belief that Mime is well-meaning is the effect Mime wants to create and not Mime's intention, which is for Siegfried to believe that Mime is well-meaning. Instead of using the labels

"lying" and "not lying" or "true" and "false," it may be beneficial to think in terms of intention and effect, not least in cases involving more complicated types of incongruity. For instance, in "Willkommen, Siegfried," instead of saying Mime is "lying" on Coating 2 and "not lying" on Coating 3—which can be rather counter-intuitive since Mime himself does not have access to Coating 3—it would make more sense to simply see the meaning on Coating 2 as the effect Mime wishes to have on Siegfried, with which he plans to achieve his intention, namely the meaning on Coating 3.

A final distinction vital to the analysis of multimodality is the distinction between the author's and the character's intention (or effect), as the coexistence of simultaneously represented information grants the author the opportunity to more freely incorporate and therefore manipulate multiple coatings of information. This is particularly obvious in Wagner's work, largely due to the highly tangible semantics of his leitmotifs. When a character says something in words, the accompanying leitmotif can conveniently communicate something else, including something the character is not consciously aware of or simply has no knowledge of. In "Als zullendes Kind zog ich dich auf," for instance, Mime has no clue the leitmotif he is singing his words to is giving him away to the audience. Likewise, when a character's music is already "saying" what the character intends to convey, the character's words can say otherwise, including things the character has no idea about, as is the case in "Willkommen, Siegfried," where Mime's music "says" what he intends to say to Siegfried, but where

he has no idea what his speech actually sounds like in Siegfried's—and therefore the listener's—ears. In cases like these, not all information conveyed by a character is within the character's intention. Rather, some information could be something that the character has no control over or is not aware of, something that the author intends to convey through the character for reasons such as composition, structure, progression of plot, or the like. Arrangements like these are possible only because of layering.

Whereas the cognitively oriented analysis in Chapter 4 has captured the generalities of multimodal meaning construction in time in the two Wagner passages, in the present chapter I have discussed in further detail the idiosyncrasies of these two opera excerpts, from a primarily social-interactional perspective. The peculiarities of the usage events in the two passages seem to be attributable to the fact that they take place in a fictional context, namely Wagner's music drama *der Ring des Nibelungen*. This allows for situations that are impossible in the real world, such as the supernatural power of the dragon blood and the ability to understand the language of nature. Another factor contributing to these peculiarities is that the usage events are multimodal. The simultaneous presence of multiple channels allows not just complementary but also contrasting pieces of information to be communicated at the same time, a fact which provides the composer with additional means of achieving compositional effects, and which allows the listener to observe the multiple coatings of meaning more clearly and easily. Importantly, the stark contrast between the ease of

our understanding of the two Wagner passages and the complexity of the above discussion shows the agility of the cognitive processing that allows us to quickly make sense of the world we are exposed to.

# 6. Conclusion and future directions.

Though ubiquitous in all multimodal means of communication and essential to the understanding of the most ordinary, the temporally simultaneous relation of expectation, unlike the temporally sequential anticipation, has hitherto been largely overlooked in the literature. To bridge the gap, in the present study I have identified theoretical notions central to the analysis of phenomena involving expectation—including the progression of time, conceptual distances, the processes of conceptual integration and disintegration, constructions, intention, etc.—and put together a unified framework that is capable of capturing the nuances and dynamics of multimodal meaning construction. As shown by the analysis of the two *Siegfried* passages, which have very different language-music dynamics, cross-modal mismatch brings about drastic changes in conceptual distances, imposing tension on the listener's expectation and prompting the listener to come up with emergent interpretations through conceptual integration and disintegration. The analysis has also demonstrated that, much like unfulfilled anticipation, unfulfilled expectation also yields strong emotional effects. Though seemingly less direct than cases where language and music match and where conceptual distances are short, in passages with cross-modal unexpectedness, the clash between incompatible meaningful structures often contributes to especially rich meaning and emotional effects on the listener, which may be what the composer intends to achieve by composing in an

"incongruous," "non-literal" way in the first place. Turning to the social-interactional aspect of the phenomenon, I have further tackled the residual idiosyncrasies of the two *Siegfried* passages. By breaking down the multiplex meanings associated with Mime into multiple layers and coatings, the actual complexity of the passages and the cognitive processing involved has been further laid out, not least that of "Willkommen, Siegfried."

While theoretical framework of the present study is presented as a theoretical toolbox, it is much more than an amalgam of theories. Rather, it draws upon insightful notions from numerous relevant fields, incorporating them in a way that is compatible with multimodal research, and offering a potentially common platform on which further multimodal research can be based. For instance, the incorporation of conceptual distances has provided a consistent way to measure degrees of functional similarity in between multimodal constructions; using construction as a cross-modally valid unit of analysis has also facilitated comparisons across different modalities. Importantly, accommodating the complexity of the intertwining cognitive operations involved in multimodal means of communication, the toolbox as a general framework is not confined to handling cases with cross-modal mismatch, but also captures the nuances of cross-modal dynamics even where incongruity is absent, as the analysis of the two *Siegfried* passages shows. In addition to the contributions of the theoretical toolbox, the identification of coatings in meaning communication also complements

existing research along the line of social-interactional linguistics, deepening and broadening our understanding of complex cases of language use.

Since the relation of expectation exists by definition in all kinds of multimodal communication where simultaneous events are possible, the discoveries and results of the present study also shed light on research in other relevant fields, such as gesture-speech relations, simultaneous constructions in signed language, movies, as well as chamber music rehearsal and orchestral conducting. Multimodal phenomena often considered monomodal, such as verbal irony, humor, and sarcasm, all of which have long puzzled linguists, can likewise benefit from the findings of the present research.

Also awaiting further investigation is empirical testing of the toolbox and other proposals made in the present research, such as the actual measuring of conceptual distances, whether constructions are indeed the building blocks of multimodal events, the online processing of conceptual integration and disintegration. The social-interactional aspect of multimodal meaning construction can also benefit from further investigation, particularly in the empirical testing of the cognitive processing of coatings of meaning. With support from empirical research, the proposals in the present study will then be able to be modified accordingly, and become more comprehensive and well-rounded.

# Appendix

**MIME**
Willkommen, Siegfried!
Sag', du Kühner,
hast du das Fürchten gelernt?

**SIEGFRIED**
Den Lehrer fand ich noch nicht.

**MIME**
Doch den Schlangenwurm,
du hast ihn erschlagen:
das war doch ein schlimmer Gesell?

**SIEGFRIED**
So grimm und tückisch er war,
sein Tod grämt mich doch schier,
da viel üblere Schächer
unerschlagen noch leben!
Der mich ihn morden hieß,
den hass' ich mehr als den Wurm.

**MIME**
Nur sachte! Nicht lange
sieh'st du mich mehr:
zum ew'gen Schlaf
schließ' ich dir die Augen bald!
Wozu ich dich brauchte,
hast du vollbracht;
jetzt will ich nur noch
die Beute dir abgewinnen: —
mich dünkt, das soll mir gelingen;
zu bethören bist du ja leicht!

**SIEGFRIED**
So sinn'st du auf meinen Schaden?

**MIME**
Welcome, Siegfried!
Tell me, brave boy,
have you learned the meaning of fear?

**SIEGFRIED**
I've not yet found a teacher.

**MIME**
But you've slain
the snake-like dragon:
he must have been a poor companion?

**SIEGFRIED**
Grim and spiteful though he was,
his death yet grieves me deeply
since far worse villains
still remain unslain!
The man who bade me murder him
I hate much more than the dragon.

**MIME**
But soft! You'll not
have to see me much longer:
I'll soon lock
your eyes in lasting sleep!
You've done
what I needed you for;
all that I still want to do
is to win from you the booty: —
I think that I ought to succeed in that;
you're easy enough to fool after all!

**SIEGFRIED**
So you're planning to do me harm?

**MIME**

Wie sagt' ich denn das?
Siegfried, hör' doch, mein Söhnchen!
Dich und deine Art
haßt' ich immer von Herzen;
aus Liebe erzog ich
dich lästigen nicht:
dem Horte in Fafner's Hut,
dem Golde galt meine Müh'.
Gibst du mir das
gutwillig nun nicht: —
Siegfried, mein Sohn,
das sieh'st du wohl selbst —
dein Leben mußt du mir lassen!

**SIEGFRIED**

Daß du mich hassest,
hör' ich gern:
doch auch mein Leben muß ich dir lassen?

**MIME**

Das sagt' ich doch nicht?
Du versteh'st mich ja falsch!
Sieh', du bist müde
von harter Müh';
brünstig wohl brennt dir der Leib:
dich zu erquicken
mit queckem Trank
säumt' ich Sorgender nicht.
Als dein Schwert du dir branntest,
braut' ich den Sud:
trink'st du nun den,
gewinn' ich dein trautes Schwert,
und mit ihm Helm und Hort.

**SIEGFRIED**

So willst du mein Schwert
und was ich erschwungen,
Ring und Beute mir rauben?

**MIME**

What, did I say that?
Siegfried, sonny, listen to me!
You and your kind
I have always hated with all my heart;
it was not out of love
that I brought you up, you burdensome child:
my efforts were aimed at the gold,
at the hoard in Fafner's safekeeping.
If you don't give it up
to me willingly now,
Siegfried, my son,
you must see for yourself —
you must yield up your very life to me!

**SIEGFRIED**

That you hate me
I'm glad to hear:
but must I also yield up my life?

**MIME**

Surely I didn't say that?
You've understood me all wrong!
Look, you are tired
from toilsome exertions;
your body must burn with a raging fire:
anxious as ever, I didn't delay
but came to restore you
with quickening draught.
While you were smelting your sword,
I was brewing this broth:
if you'll drink it now,
I'll win your trusty sword
and with it helmet and hoard.

**SIEGFRIED**

So you'd steal my sword
and all I have won
by way of ring and booty?

**MIME**

Was du doch falsch mich versteh'st!
Stamml' ich, fasl' ich wohl gar?
Die größte Mühe
geb' ich mir doch,
mein heimliches Sinnen
heuchelnd zu bergen,
und du dummer Bube
deutest alles doch falsch!
Öffne die Ohren
und vernimm genau:
höre, was Mime meint! —
Hier nimm, und trinke die Labung!
Mein Trank labte dich oft:
that'st du auch unwirsch,
stelltest dich arg:
was ich dir bot —
erbos't auch—nahm'st du's doch immer.

**MIME**

How you mistake my meaning!
Do I stammer or even talk rubbish?
I'm taking the greatest
pains after all
to hide, by dissembling,
my secret thoughts
and you, stupid boy,
interpret everything wrongly!
Open your ears
and listen closely:
hear what Mime means! —
Take this, and drink this refreshment!
My drink has often refreshed you:
though you acted morosely,
affecting malice,
you always took
what I offered, even when enraged.

**SIEGFRIED**

Einen guten Trank
hätt' ich gern:
wie hast du diesen gebrau't?

**SIEGFRIED**

I'd be glad
of something good to drink:
how did you brew this one here?

**MIME**

Hei, so trink' nur:
trau' meiner Kunst!
In Nacht und Nebel
sinken die Sinne dir bald:
ohne Wach' und Wissen
stracks streck'st du die Glieder.
Lieg'st du nun da,
leicht könnt' ich
die Beute nehmen und bergen:
doch erwachtest du je,
nirgends wär' ich
sicher vor dir,
hätt' ich selbst auch den Ring.

**MIME**

Hey, just drink it:
trust in my art!
In darkness and mist
your senses will soon be shrouded:
unwaking, unwitting,
you'll straightway stretch out your limbs.
And once you're lying there,
then I could easily
take the spoils and conceal them:
but if you were ever to waken,
nowhere would I be
safe from you,
though I had the ring itself.

D'rum mit dem Schwert,
das so scharf du schuf'st,
hau' ich dem Kind
den Kopf erst ab:
dann hab' ich mir Ruh'
und auch den Ring!

**SIEGFRIED**
Im Schlafe willst du mich morden?

**MIME**
Was möcht' ich? Sagt' ich denn das? —
Ich will dem Kind
nur den Kopf abhau'n.
Denn haßte ich dich
auch nicht so sehr,
und hätt' ich des Schimpf's
und der schändlichen Mühe
auch nicht so viel zu rächen:
aus dem Wege dich zu räumen
darf ich doch nicht rasten,
wie käm' ich sonst anders zur Beute,
da Alberich auch nach ihr lugt? — —
Nun, mein Wälsung!
Wolfssohn du!
Sauf' und würg' dich zu todt:
nie thu'st du mehr 'nen Schluck!

**SIEGFRIED**
Schmeck' du mein Schwert,
ekliger Schwätzer!

And so with the sword
which you made so sharp
I'll first hack off
the child's head:
then I'll have peace of mind
and the ring as well!

**SIEGFRIED**
You mean to murder me while I'm asleep?

**MIME**
I mean to do what? Is that what I said? —
I want only
to hack the child's head off.
For even if
I hated you less
and hadn't so much
of your hateful abuse
and such shameful toil to avenge,
I'd still waste no time
in clearing you out of the way
for how else could I gain the spoils,
since Alberich covets them, too? — —
Now, my Wälsung!
Son of Wolfe!
Drink and choke yourself to death:
you'll never taste another drop!

**SIEGFRIED**
Have a taste of my sword,
you loathsome babbler!

# REFERENCES


Aikhenvald, Alexandra Y. 2012. The essence of mirativity. Linguistic Typology 16(3), 435–86.

Arnheim, Rudolf. 1974. *Art and Visual Perception: A Psychology of the Creative Eye*, the new edition. Berkeley, CA: University of California Press.

Attardo, Salvatore. 1997. The semantic foundations of cognitive theories of humor. *Humor: International Journal of Humor Research* 10(4), 395–420.

Attardo, Salvatore. 2000. Irony as relevant inappropriateness. *Journal of Pragmatics* 32, 793–826.

Attardo, Salvatore, Eisterhold, Jodi, Hay Jennifer, and Poggi, Isabella. 2003. Multimodal markers of irony and sarcasm. *Humor* 16(2), 243–60.

Attardo, Salvatore, Hempelmann, Christian F., and Di Maio, Sara. 2002. Script oppositions and logical mechanisms: Modeling incongruities and their resolutions. *Humor: International Journal of Humor Research* 15(1), 3–46.

Bache, Carl. 2005. Constraining conceptual integration theory: Levels of blending and disintegration. *Journal of Pragmatics* 37, 1615–35.

Barthes, Roland, 1964/1977. *Image, Music, Text*, trans. by Heath, Stephen. New York: Noonday.

Bashwiner, David. 2013. Music analysis for multimedia: A perspective from music theory. In Tan, Siu-Lan, Cohen, Annabel J., Lipscomb, Scott D., and Kendall,

Roger A. (Eds.), *The Psychology of Music in Multimedia*, 89–117. Oxford: Oxford University Press.

Bergen, Benjamin, and Chang, Nancy. 2013. Embodied Construction Grammar. In Hoffman, Thomas, and Trousdale, Graeme (Eds.), *The Oxford Handbook of Construction Grammar*, 168–90. Oxford: Oxford University Press.

Bernstein, Leonard. 1976. *The Unanswered Question: Six Talks at Harvard*. Cambridge, MA: Harvard University Press.

Bolinger, Dwight. 1965. *Forms of English: Accent, Morpheme, Order*, ed. by Abe, Isamu, and Kanekiyo, Tetsuya. Cambridge, MA: Harvard University Press.

Bolinger, Dwight. 1985. *Intonation and Its Parts: Melody in Spoken English*. London: Edward Arnold.

Bolinger, Dwight. 1989. *Intonation and its Uses: Melody in Grammar and Discourse*. London: Edward Arnold.

Bolivar, Valerie J., Cohen, Annabel J., and Fentress, John C. 1994. Semantic and formal congruency in music and motion pictures: Effects on the interpretation of visual action. *Psychomusicology* 13, 28–59.

Boltz, Marilyn. 2004. The cognitive processing of film and musical soundtracks. *Memory & Cognition* 32, 1194–205.

Boltz, Marilyn, Schulkind, Matthew, and Kantra, Suzanne. 1991. Effects of background music on the remembering of filmed events. *Memory & Cognition*, 19(6), 593–606.

Bratman, Michael E. 1987. *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

Bratman, Michael E. 1997. I intend that we J. In Tuomela, R., and Holmstrom-Hintikka, G. (Eds.), *Contemporary Action Theory, Vol. 2: Social Action*, 49–63. Dordrecht: Kluwer.

Brône, Geert. 2012. Humour and irony in cognitive pragmatics. In Schmid, Hans-Jörg (Ed.), *Cognitive Pragmatics*, 463–504. Berlin: De Gruyter Mouton.

Chafe, Wallace. 1977a. Creativity in verbalization and its implications for the nature of stored knowledge. In Freedle, R. (Ed.), *Discourse Production and Comprehension*, 41–55. Norwood, NJ: Ablex.

Chafe, Wallace. 1977b. The recall and verbalization of past experience. In Cole, Peter. (Ed.), *Current Issues in Linguistic Theory*, 215–46. Bloomington: Indiana University Press.

Chua, Daniel K. L. 2007. Rioting with Stravinsky: A particular analysis of the *Rite of Spring*. *Music Analysis* 26(i-ii), 59–109.

Clark, Herbert H. 1996. *Using Language*. Cambridge: Cambridge University Press.

Cohen, Annabel J. 2010. Music as a source of emotion in film. In Juslin, Patrik N., and Sloboda, John A. (Eds.), *Handbook of Music and Emotion: Theory, Research, Applications*, 879–908. Oxford: Oxford University Press.

Cohen, Annabel, J. 2013. Congruence-Association Model of music and multimedia: Origin and evolution. In Tan, Siu-Lan, Cohen, Annabel J., Lipscomb, Scott D., and Kendall, Roger A. (Eds.), *The Psychology of Music in Multimedia*, 18–47. Oxford: Oxford University Press.

Cook, Nicholas. 1998. *Analysing Musical Multimedia*. Oxford: Oxford University Press.

Coulson, Seana. 2001. *Semantic Leaps: Frame-Shifting and Conceptual Blending in Meaning Construction*. Cambridge: Cambridge University Press.

Coulson, Seana. 2005. Sarcasm and the Space Structuring Model. In Coulson, Seana, Lewandowska-Tomaszczyk, Barbara (Eds.), *The Literal and the Nonliteral in Language and Thought*, 129–144. Berlin: Peter Lang.

Coulson, Seana, King, Jonathan W., and Kutas, Marta. 1998. Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes* 13(1), 21–58.

Coulson, Seana, and Kutas, Marta. 2001. Getting it: Human event-related brain response to jokes in good and poor comprehenders. *Neuroscience Letter* 316, 71–74.

Craft, Robert. 1966. *The Rite of Spring*: Genesis of a masterpiece. *Perspectives of New Music* 5(1), 20–36.

Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Croft, William. 2003. *Typology and Universals*, 2nd Ed. Cambridge: Cambridge University Press.

Croft, William. 2007. The origins of grammar in the verbalization of experience. *Cognitive Linguistics* 18(3), 339–82.

Croft, William. 2009. Toward a social cognitive linguistics. In Evans, Vyvyan, and Pourcel, Stéphanie (Eds.), *New Directions in Cognitive Linguistics*, 395–420. Amsterdam: John Benjamins.

Croft, William. 2011. Language structure in its human context: New directions for the language sciences in the twenty-first century. In Hogan, Patrick Colm (Ed.), *The Cambridge Encyclopedia of the Language Sciences*, 1–10. Cambridge: Cambridge University Press.

Croft, William, and Cruse, D. Alan. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Davenport, Tristan, Coulson, Seana. 2011. Predictability and novelty in literal language comprehension: An ERP study. *Brain Research* 51(5), 907–21.

DeLancey, Scott. 1997. Mirativity: The grammatical marking of unexpected information. *Linguistic Typology* 1, 32–52.

Donington, Robert, 1974. *Wagner's 'Ring' and Its Symbols*, 3rd Ed. London: Faber and Faber.

Du Bois, John W. 2007. The stance triangle. In Englebretson, Robert (Ed.), *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*, 139–82. Amsterdam: John Benjamins.

Du Bois, John W. 2014. Towards a dialogic syntax. *Cognitive Linguistics* 25(3), 359–410.

Fauconnier, Gilles, and Turner, Mark. 2002. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: BasicBooks.

Festinger, Leon. 1957. *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.

Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di Semantica* 6, 222–55.

Fillmore, Charles J., and Atkins, Beryl T. 1992. Toward a frame-based lexicon: The semantics of RISK and its neighbors. In Lehrer, Adrienne, and Kittay, Eva Feder (Eds.), *Frames, Fields, and Contrasts: New Essays in Semantic and Lexical Organization*, 75—102. Hillsdale, NJ: Lawrence Erlbaum Associates.

Fillmore, Charles J., Kay, Paul, and O'Connor, Mary Catherine. 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64, 501–38.

Gärdenfors, Peter. 2000. *Conceptual Spaces: The Geometry of Thought*. Cambridge, MA: MIT Press.

Gärdenfors, Peter. 2014. *The Geometry of Meaning*. Cambridge, MA: MIT Press.

Geeraerts, Dirk, and Cuyckens, Hubert. 2007. Introduction to Cognitive Linguistics. In Geeraerts, Dirk, and Cuyckens, Hubert (Eds.), *The Oxford Handbook of Cognitive Linguistics*, 3–21. Oxford: Oxford University Press.

Gibbs, Raymond W. Jr. 2012. Are ironic acts deliberate? *Journal of Pragmatics* 44, 104–15.

Giora, Rachel. 2003. *On Our Mind: Salience, Context and Figurative Language*. Oxford: Oxford University Press.

Gjerdingen, Robert, and Bourne, Janet. 2015. Schema theory as construction grammar. *Music Theory Online* 21(2).

Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Grice, H. Paul. 1957. Meaning. *The Philosophical Review* 66(3), 377–88.

Haiman, John. 1983. Iconic and economic motivation. *Language* 59(4), 781–819.

Herbert, James D. 2002. The debts of divine music in Wagner's *Ring des Nibelungen*. *Critical Inquiry* 28(3), 677–708.

Herbert, James D. 2008. *Our Distance From God: Studies of the Divine and the Mundane in Western Art and Music*. Berkeley, CA: University of California Press.

Hougaard, Anders. 2002. Conceptual disintegration as a principle of meaning construction. Pre-publications of the English Department of Odense University, The University of Southern Denmark, Issue 119.

Hougaard, Anders. 2005. Conceptual disintegration and blending in interactional sequences: A discussion of new phenomena, processes vs. products, and methodology. *Journal of Pragmatics* 37, 1653–85.

Hsu, Hui-Chieh, and Su, Lily I-wen. 2014. Love in disguise: Incongruity between text and music in song. *Journal of Pragmatics* 62, 136–50.

Huron, David. 2006. *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.

Kemmer, Suzanne. 1993. *The Middle Voice*. Amsterdam: John Benjamins.

Kelly, Spencer D., Özyürek, Aslı, and Maric, Eric. 2010. Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science* 21(2), 260–7.

Kelly, Spencer D., Kravitz, Corinne, and Hopkins, Michael. 2004. Neural correlates of bimodal speech and gesture comprehension. *Brain and Language* 89, 253–60.

Kelly, Spencer D., Ward, Sarah, Creigh, Peter, and Bartolotti, James. 2007. An intentional stance modulates the integration of gesture and speech during comprehension. *Brain and Language* 101, 222–33.

Kramer, Lawrence. 1990. Culture and musical hermeneutics: The Salome complex. *Cambridge Opera Journal*, 2(3), 269–94.

Kreuz, Roger J. and Roberts, Richard M. 1995. Two cues for verbal irony: Hyperbole and the ironic tone of voice. *Metaphor and Symbolic Activity* 10, 21–31.

Lakoff, George, and Johnson, Mark. 1980. *Metaphors We Live by*. Chicago: Chicago University Press.

Levitin, Daniel J. 2006. *This is Your Brain on Music: The Science of a Human Obsession*. New York: Plume.

Lewis, David K. 1969. *Convention*. Cambridge, MA: Harvard University Press.

Magee, Bryan. 2000. *The Tristan Chord: Wagner and Philosophy*. New York: Henry Holt.

Marshall, Sandra K., and Cohen, Annabel J. 1988. Effects of musical soundtracks on attitudes toward animated geometric figures. *Music Perception: An Interdisciplinary Journal* 6(1), 95–112.

Meyer, Leonard B. 1956. *Emotion and Meaning in Music*. Chicago: Chicago University Press.

Michaelis, Laura. 2013. Sign-based Construction Grammar. In Hoffman, Thomas, and Trousdale, Graeme (Eds.), *The Oxford Handbook of Construction Grammar*, 133–52. Oxford: Oxford University Press.

Millington, Barry (Ed.). 2006. *The New Grove Guide to Wagner and His Operas*. Oxford: Oxford University Press.

Monahan, Seth. 2007. "Inescapable" coherence and the failure of the novel-symphony in the Finale of Mahler's Sixth. *19th-Century Music* 31(1), 53–95.

Newman, Ernest. 1949. *Wagner Nights*. London: The Bodley Head.

Puffett, Derrick (Ed.). 1989. *Richard Strauss: Salome*. Cambridge: Cambridge University Press.

Reisenzein, Rainer, Meyer, Wulf-Uwe., and Niepel, Michael. 2012. Surprise. In Ramachandran, V.S. (Ed.), *Encyclopedia of Human Behavior*, 2nd Ed. London: Academic Press.

Rosch, Eleanor, and Lloyd, Barbara (Eds). 1978. *Cognition and Categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Russell, James A. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39(6), 1161–78.

Sabor, Rudolph. 1997a. *Richard Wagner,* Der Ring des Nibelungen*: A Companion Volume*. London: Phaidon.

Sabor, Rudolph. 1997b. *Richard Wagner,* Siegfried*: Translation and Commentary*. London: Phaidon.

Shevy, Mark. 2007. The mood of rock music affects evaluation of video elements differing in valence and dominance. *Psychomusicology* 19(2), 57–78.

Spencer, Stewart, and Millington, Barry (Eds.). 1993. *Wagner's Ring of the Nibelung: A Companion*. New York: Thames and Hudson.

Sperber, Dan and Wilson, Deirdre. 1981. Irony and the use-mention distinction. In Cole, Peter. (Ed.), *Radical Pragmatics*, 295–318. New York: Academic Press.

St. George, Marie, Mannes, Suzanne, and Hoffman, James E. 1994. Global semantic expectancy and language comprehension. *Journal of Cognitive Neuroscience* 61(1), 70–83.

Stassen, Leon. 1997. *Intransitive Predication*. Oxford: Oxford University Press.

Steels, Luc. 2013. Fluid Construction Grammar. In Hoffman, Thomas, and Trousdale, Graeme (Eds.), *The Oxford Handbook of Construction Grammar*, 153–67. Oxford: Oxford University Press.

Teng, Norman Y., and Sun, Sewen, 2002. Grouping, simile, and oxymoron in pictures: A design-based cognitive approach. *Metaphor and Symbol* 17, 295–316.

Tomasello, Michael. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Tsai, Chen-Gia, Chen, Rong-Shan, and Tsai, Tzung-Shian. 2014. The arousing and cathartic effects of popular heartbreak songs as revealed in the physiological responses of listeners. *Musicae Scientiae* 18(4), 410–22.

Van Overwalle, Frank, and Jordens, Karen. 2002. An adaptive connectionist model of cognitive dissonance. *Personality and Social Psychology Review* 6, 204–31.

Veale, Tony, Feyaerts, Kurt, and Forceville, Charles. 2013. E unis pluribum: Using mental agility to achieve creative duality in word, image and sound. In Veale, Tony, Feyaerts, Kurt, and Forceville, Charles (Eds.), *Creativity and the Agile Mind: A Multi-Disciplinary Study of a Multi-Faceted Phenomenon*, 37–57. Berlin: De Gruyter Mouton.

Vermeerbergen, Myriam, Leeson, Lorraine, and Crasborn, Onno Alex (Eds.). 2007. *Simultaneity in Signed Languages: Form and Function*. Amsterdam: John Benjamins.

Vitouch, Oliver. 2001. When your ear sets the stage: Musical context effects in film perception. *Psychology of Music* 29, 70–83.

Wilcox, Sherman, and Shaffer, Barbara. 2006. Modality in American Sign Language. In Frawley, William (Ed.), *The Expression of Modality*, 207–238. Berlin: De Gruyter Mouton.

Willemsen, Steven, and Kiss, Miklós. 2013. Unsettling melodies: A cognitive approach to incongruent film music. *Film and Media Studies* 7, 169–83.

Wu, Ying Choon, and Coulson, Seana. 2005. Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology* 42, 1–15.

Wu, Ying Choon, and Coulson, Seana. 2007. How iconic gestures enhance communication: An ERP study. *Brain and Language* 101, 234–45.

Zbikowski, Lawrence M. 1999. The blossoms of 'Trockne Blumen': Music and text in the early nineteenth century. *Music Analysis* 18, 307–45.

Zbikowski, Lawrence M. 2002. *Conceptualizing Music: Cognitive Structure, Theory, and Analysis*. Oxford: Oxford University Press.

Zbikowski, Lawrence M. 2011. Musical gesture and musical grammar: A cognitive approach. In Gritten, Anthony, and King, Elaine (Eds.), *New Perspectives on Music and Gesture*, 83–98.

# Data

Wagner, Richard. 1871. *Siegfried*, WWV 86C.