Fachbereich Erziehungswissenschaft und Psychologie

der Freien Universität Berlin

Affective Iconicity in Language and Poetry

A Neurocognitive Approach

DISSERTATION

zur Erlangung des akademischen Grades

Doktor der Philosophie (Dr. phil.)

Doctor of Philosophy (Ph.D.)

Vorgelegt von

Arash Aryani (M.Sc./B.Sc.)

Berlin, 2018

Erstgutachter:

Univ.-Prof. Dr. Arthur M. Jacobs (Freie Universität Berlin)

Zweitgutachter:

Univ.-Prof. Dr. Hauke Heekeren (Freie Universität Berlin)

Disputation:

12. July 2018

Acknowledgment

Writing this note of thanks is the finishing touch to my dissertation. It has been a period of intense learning for me in both the scientific arena and on a personal level. I would like to reflect on the people who have supported and helped me so much throughout this period.

Foremost, I would like to thank my supervisor Arthur Jacobs for supporting me through this spectacular journey of interdisciplinary research. Arthur, thank you for trusting me and for motivating me from the first day that I came to your office after I quit my job in industry to pursue my passion in the neuroscience of language. You always made me believe in myself and taught me what it takes to become a good scientist.

I am very grateful to Hauke Heekeren for agreeing to be the second reviewer of my dissertation despite his busy schedule.

For creating a stimulating and friendly scientific environment, I would like to thank my former and current colleagues and friends from Arthur Jacobs' lab at the Freie Universität Berlin: Eva Fröhlich, Chun-Ting Hsu, Michael Kuhlmann, Johanna Liebig, Jana Lüdtke, Gesche Schauenburg, David Schmidtke, Teresa Sylvester, Sascha Tamm, Susann Ullrich, and Shuwei Xue.

I would like to express my gratitude to my former supervisor Markus Conrad for his guidance and mentorship during my first steps, and for the intriguing discussions that contributed significantly to the direction of this dissertation.

Finally, I would like to thank my friends and family for their support, in particular my parents for always encouraging me to pursue my passion and go my own way, Amir for supporting me and being a role model in many aspects of my life, Omid for being a generous brother and a great friend, Ida for her love, her generosity, and her support, Jenny for pushing me to strive towards my goals and for supporting me through all the difficulties, Ed and Maria for their great help with proofreading, and all my Berlin friends who have made the city home for me.

Summary

One of the most important hallmarks of human language is the use of symbolic signs in that words just “symbolically” represent the objects they refer to. That is, neither the written word ‘tree’ nor the spoken sequence of sounds /triː/ has any inherent similarity or analogy to a real tree. This is a core assumption of classic linguistics, which states that words are paired with objects, mental images, or concepts in an arbitrary fashion. Thus, the sound of a word per se has no inherent semantic content, nor does it play any contributing role in shaping the meaning of words.

The goal of this dissertation is to contemplate, to rethink, and to examine the aforementioned statement, which has dominated language research throughout the last century. Indeed, there are a vast number of counterexamples showing how meaningful single phonemes, or their combination—as in nonwords—can be. Consider the standalone role of sound in poetry, or the use of single syllables or phonemes in various sacred rituals, or the prevalence of onomatopoetic words, i.e., words that sound similar to what they mean (e.g., click, zigzag), across all of the languages in the world, or the tendency of language users toward using harsh-sounding words as swear words, or cross-linguistic phenomena such as a preference to match the nonword BOUBA with a curvy round shape and KIKI with a spiky angular shape: These all constitute excellent examples that could potentially falsify the radical assumption of the sound of words being per se meaningless.

The present dissertation now wants to shed new empirical light on this old debate, which dates back to Greek antiquity, yet faces a number of unanswered questions regarding the cognitive and neuropsychological mechanisms underlying the potential effect of sound on the processes of meaning making. More specifically, my focus is on the existence of sound-meaning relationships in the affective domain, termed *affective iconicity*, and on the investigation of different aspects of this phenomenon in both everyday language and poetry. By taking an interdisciplinary approach, the present work combines a variety of methods and techniques, such as behavioral and neuroimaging experiments, phonological and acoustic analysis of a

large-scale lexicon, computational modelling, and corpus analysis, in order to provide comprehensive answers for this multi-faceted phenomenon.

The theoretical part of the dissertation explores and integrates linguistic perspectives on the *iconic* mappings, explaining how a linguistic sign can acquire meaning based on similarities between its form and the object it refers to. A surprising neglect of the role of emotion in empirical models of language in general, and in previous investigations of iconicity in particular, is discussed. Specific hypotheses are formulated via the predictions made by the recently proposed Neurocognitive Poetics Model (NCPM), and through reviewing the previous works on the topic. Accordingly, three main questions are formulated that the present dissertation aims to address: i) Does the sound of words evoke affective responses observable at the behavioral and neural level? ii) Does the sound of words influence the processes of meaning making in the affective domain? iii) Does the sound of words in a poem contribute to its global affective meaning as perceived by readers? Six empirical studies attempt to address these questions which are subdivided into six more precise research questions.

Results of the empirical part provide a comprehensive picture of the interplay between sound and meaning at different levels of processing (i.e., rating, semantic decision, and passive listening) for different presentation modalities (i.e., visual, and auditory) and for different textual levels (i.e., single word, and entire text). In short, results of Study 1 and Study 2 indicated a high similarity between the affective potential of the sound of words and other types of affective sounds (e.g., nonverbal emotional vocalization and affective prosody) at both the level of psychological perception (Study 1) and the level of neural correlates and substrates (Study 2). Furthermore, when giving their affective judgments (valence and arousal) about the meaning of words, participants, as shown in Study 1, were implicitly influenced by the sound of words even when words were presented visually and read silently. These results were extended in Study 3 in which iconic words, as operationalized by congruence between affective sound and affective meaning, were evaluated more quickly and more accurately than their non-iconic counterparts, suggesting that a similarity between the form and meaning of a word may help language users to more readily access its meaning through direct form-meaning mappings. Study 4 investigated the neural mechanisms underlying the facilitative effect observed in Study 3. Results showed an enhanced fMRI signal in the left amygdala, known for its role in multimodal emotion integration, for both a comparison between iconic and non-iconic words, as well as functional connectivity between two seed regions representing the sound (superior temporal gyrus) and meaning (inferior frontal gyrus) of words modulated by iconic condition. Lastly, results of Study 5 and Study 6 emphasize the role of foregrounded phonological units in the affective and aesthetic processes of literary reading. This clearly supports the initial hypotheses that iconicity is a feasible indicator of the affective qualities of a literary text as evoked by particular phonemic structures. The presented method for measuring the basic affective tone of the poems investigated could account for a considerable part of the variance in the ratings of their general affective meaning.

In summary, this dissertation provides strong psychological and neuroimaging evidence for a device that has long been deployed in poetry and the arts, i.e., evoking affective (and aesthetic) responses by the use of certain words with specific sound patterns. The results were used to upgrade the standard models of visual word processing by conceiving corresponding modules responsible for the evaluation of affective sound and its interaction with the evaluation of the affective meaning of words. Lastly, at the more complex level of the whole text, the findings of this dissertation confirm the central assumption of the NCPM regarding the role of foregrounded elements in enhancing affective perception, although the Panksepp-Jakopson hypothesis might need to be extended to human-specific brain regions which originally evolved for other, more simple, tasks. Also, the literary model of reading may need to be updated by adding feedback loops from resulting reading behavior (e.g., fluent reading) to the perceived emotions (e.g., lust and play) based on the findings concerning the facilitative role of iconicity in language processing.

Zusammenfassung

Eine der wichtigsten Eigenschaften der menschlichen Sprache ist die Verwendung symbolischer Zeichen, in der Wörter die Objekte, auf die sie referieren, nur „symbolisch“ repräsentieren: Weder das geschriebene Wort „Baum“ noch die gesprochene Sequenz der Laute /bau̯m/ weisen Ähnlichkeiten mit einem realen Baum auf. In der klassischen Linguistik wird angenommen, dass die Beziehung zwischen Wörtern und Objekten völlig arbiträr ist. Das heißt, dass der Klang von Wörtern keinerlei eigenes semantisches Gewicht hat und somit keine Rolle für die Bedeutung eines Wortes spielt.

Das Ziel dieser Dissertation ist es, diese Annahme, die die linguistischen Theorien in den letzten Jahrzenten geprägt hat, empirisch zu überprüfen und zu untersuchen. Es gibt zahlreiche Gegenbeispiele, die darauf hinweisen, dass einzelne Phoneme oder Kombinationen von Phonemen durchaus eine eigene Bedeutung inhärieren können: Die Rolle klanglicher Strukturen in der Poesie, die Verwendung einzelner Silben oder Phoneme in Ritualen, der Einsatz sogenannter onomatopoetischer Wörter (z.B. klick) über Sprachen hinweg, oder die Assoziation des Pseudowortes KIKI mit kantigen Formen und BOUBA mit runden-kurvigen Formen, sind nur einige Beispiele, die eine arbiträre Beziehung zwischen Laut und Bedeutung der Wörter in Frage stellen.

In der vorliegenden Dissertation werden diese Fragen, die bereits in der griechischen Antike aufgeworfen wurden und auch heute noch debattiert werden, bearbeitet. Der Fokus liegt dabei auf den kognitiven und neuropsychologischen Mechanismen, die einer Klang-Bedeutung in der Interaktion zugrunde liegen. Im Speziellen beschäftigt sich diese Arbeit mit dem Zusammenspiel von Klang und Bedeutung in der affektiven Domäne; der affektiven Ikonizität, ein Phänomen, das sowohl in der Alltagsprache als auch in der Poesie eine wichtige Rolle spielt. Es wird ein interdisziplinärer Ansatz gewählt, um das komplexe und vielschichtige Phänomen der Klang-Bedeutung-Relation zu beleuchten. Dazu werden behaviorale und

neurokognitive Experimente, phonologische und akustische Analysen sowie komputationale Modellierungen und Korpus-Analysen kombiniert.

Der theoretische Teil dieser Arbeit integriert verschiedene linguistische Perspektiven zu ikonischen Beziehungen der linguistischen Zeichen. Zudem werden die Vernachlässigung von Emotionen in empirischen Modellen zur Sprachverarbeitung und in Studien zur Relevanz von Ikonizität in der Sprache diskutiert. Basierend auf den Vorhersagen des Neurokognitven Poetischen Modells (NCPM) sowie der aktuellen Forschungslage werden spezifische Hypothesen abgeleitet. Daraus resultieren drei zentrale Fragestellungen, die in der vorliegenden Arbeit bearbeitet werden: i) Evoziert der Klang eines Wortes affektive Reaktionen, die behavioral und neuronal messbar sind? ii) Beeinflusst der Klang eines Wortes die affektive Bedeutung eines Worts? iii) Tragen die Laute der Wörter in einem Gedicht zu seiner allgemeinen affektiven Bedeutung bei? Um diese Fragen zu beantworten, wurden sechs empirische Studien durchgeführt.

Die Ergebnisse dieser empirischen Studien zeichnen ein umfassendes Bild über die Interaktion zwischen Klang und Bedeutung auf verschiedenen Ebenen der Verarbeitung (Klassifizierung, semantische Entscheidung und passives Zuhören) für verschiedene Modalitäten (visuell und auditiv) und auf verschiedenen Textebenen (einzelnes Wort, gesamter Text). Zusammengefasst deuten die Ergebnisse von Studie 1 und Studie 2 darauf hin, dass es eine hohe Ähnlichkeit zwischen dem affektiven Potenzial des Klanges von Wörtern und anderen affektiven Geräuschen (wie bspw. affektive Prosodie, non-verbale emotionale Vokalisationen) gibt. Dies kann sowohl auf der perzeptuell-psychologische Ebene (Studie 1) als auch auf der Ebene neuronaler Verarbeitung (Studie 2) gezeigt werden. Außerdem zeigt sich, dass die Entscheidung über Valenz und Arousal eines Wortes implizit durch den Klang des Wortes beeinflusst wird (Studie 1), selbst wenn dieses Wort visuell präsentiert und leise gelesen wird. Diese Ergebnisse werden in Studie 3 erweitert. Hier werden ikonische Wörter, die durch eine Kongruenz zwischen dem affektiven Klang und der affektiven Bedeutung operationalisiert werden, schneller und genauer evaluiert als deren nicht-ikonische Gegenstücke. Diese Ergebnisse weisen darauf hin, dass eine Ähnlichkeit zwischen dem

Klang eines Wortes und deren Bedeutung einen einfacheren Zugang auf die Bedeutung des Wortes ermöglichen. Studie 4 untersucht die neuronalen Mechanismen, die dem Effekt von Studie 3 zugrunde liegen könnten. Dafür wird eine verstärkte Aktivierung in der linken Amygdala, die für ihre Rolle in multimodale Integration der Emotionen bekannt ist, observiert. Diese Aktivierung gibt es sowohl für einen Vergleich zwischen ikonischen und nicht-ikonischen Wörtern, als auch in den Ergebnissen der funktionellen Konnektivität-Analyse, in der zwei Hirnareale als Kern-Regionen für die Verarbeitung des Klanges (superior temporal gyrus) und der Bededeutung (inferior frontal gyrus) berücksichtigt werden. Schließlich akzentuieren die Ergebnisse von Studie 5 und Studie 6 die Relevanz der „forgrounding“ in den affektiven und ästhetischen Prozessen des literarischen Lesens und zeigen dadurch, dass die Ikonizität ein geeigneter Indikator für die affektive Qualität eines literarischen Textes sein kann. Die vorgeschlagene Methode zur Messung des affektiven Grundtones eines Gedichtes kann einen bedeutsamen Anteil der Varianz in Bewertungen der allgemeinen affektiven Bedeutung der Gedichte aufklären.

Zusammenfassend erbringt diese Dissertation einen psychologischen und neurobiologischen Nachweis über etwas was schon längst in der Poesie und der Kunst angewendet wurde: Das Hervorrufen bestimmter affektiver und ästhetischer Reaktionen durch die Verwendung von Wörtern, die mit einer bestimmten klanglichen Eigenschaft ausgestattet sind. Die Ergebnisse dieser Studien werden weiterführend für eine Aktualisierung heutiger Modelle der Wortverarbeitung fruchtbar gemacht, indem entsprechende Module zu bestehenden Modellen hinzugefügt werden, die für die Evaluation des affektiven Klanges als auch seiner Interaktion mit der affektiven Bedeutung zuständig sind. Auf einer komplexeren Ebene der Textverarbeitung bestätigen die Befunde dieser Dissertation die zentrale Rolle der “foregrounding”. Dennoch muss die Panksepp-Jakobson-Hypothese zu menschenspezifischen Hirnarealen erweitert werden, die ursprünglich für andere Aufgaben entwickelt worden sind. Auch das Modell des literarischen Lesens soll durch das Hinzufügen der Feedbackschliefen von dem resultierendem Leseverhalten zu empfunden Emotionen aktualisiert werden.

Abbreviations

BAWL	Berlin Affective Word List
BOLD	Blood-oxygen-level dependent
CoG	Centre of Gravity
DROM	Dual Read-Out Model
ECoG	Electrocorticography
EEG	Electroencephalography
ERP	Evet related potential
fMRI	Functional Magnetic Resonance Imaging
ICC	Interclass Correlation Coefficient
IFG	Inferior Frontal Gyrus
LMM	Linear Mixed Model
MROM	Multiple Read-Out Model
NCPM	Neurocognitive Poetics Model
OP	Occipital Pole
PAP	Phonological Affective Potential
PAV	Phonological Affective Value
PCC	Posterior Cingulate Cortex
PPI	Psycho-physiological interaction
pSTS	posterior Superior Temporal Sulcus
QNA	Quantitative Narrative Analysis
SAM	Self Assistance Mankins
SAV	Sublexical Affective Value
SCN	Signal Correlated Noise
SMG	Supramarginal Gyrus
STG	Superior Temporal Gyrus
STS	Superior Temporal Sulcus
VWFA	Visual Word Form Aea

List of Original Publications

The dissertation is based on the following articles. Occasionally, the text of the original articles has been slightly modified to facilitate reading of the dissertation.

Aryani, A., Conrad, M., Schmidtke, D., & Jacobs, A. (2018). Why 'piss' is ruder than 'pee'? The role of sound in affective meaning making. *PloS one*, *13*(6), e0198430.

Aryani, A., Hsu, C. T., & Jacobs, A. M. (2018). The Sound of Words Evokes Affective Brain Responses. *Brain Sciences*, *8*(6), 94.

Aryani, A., & Jacobs, A. M. (2018). Affective Congruence between Sound and Meaning of Words Facilitates Semantic Decision. *Behavioral Sciences*, 8(6), 56.

Aryani, A., Hsu, C. T., & Jacobs, A. M. (Submitted). Affective Iconic Words Benefit from Additional Sound-Meaning Integration in the Left Amygdala

Aryani, A., Conrad, M., & Jacobs, A. M. (2013). Extracting salient sublexical units from written texts: "Emophon," a corpus-based approach to phonological iconicity. *Frontiers in Psychology*, *4*, 654.

Aryani, A., Kraxenberger, M., Ullrich, S., Jacobs, A. M., & Conrad, M. (2016). Measuring the basic affective tone of poems via phonological saliency and iconicity. *Psychology of Aesthetics, Creativity, and the Arts*, *10*(2), 191. http://dx.doi.org/10.1037/aca0000033

Table of Contents

I

Theoretical Background

Chapter 1

Introduction

Romeo Montague and Juliet Capulet, who meet and fall in love in Shakespeare's romantic tragedy are—by ill fortune—members of two warring families. In a famous quote from this story, Juliet attempts to clarify that she loves the person who is called "Montague", not the Montague name. She does so by exemplifying a rose and by emphasizing the irrelevance of verbal forms:

"What's in a name? That which we call a rose, by any other name would smell as sweet."

If a rose were a foreign flower—undocumented and nameless—we could still smell its fragrance, feel its soft petals, and be pricked by its thorns.

Juliet unintentionally alludes to the arbitrariness of the linguistic sign, a founding principle of modern linguistics, which states that the relationship between sound and meaning of words is arbitrary (Saussure, 1916/2011). Linguistic meaning, according to this notion, only emerges within a system of signs through conventions regarding the names, and hence sounds, related to certain objects. Take the word *atirgul*, for instance. The word is probably meaningless to you. Hear the word *rose*, on the other hand, and you will recall a collection of images, feelings, and associations in your mind. This doesn't happen for *atirgul*—even though it means rose in Uzbek. Thus, the principle of arbitrariness suggests that the sound of a word is per se meaningless, and in order to make sense out of the meaningless sounds you need to know the underlying conventions.

However, a number of intuitive examples from the everyday use of language and the long history of poetry suggest a possible connection between the sound of words and a layer of meaning beyond the conventional links (Jakobson, 1965; Schrott

& Jacobs, 2011). Poets often use language to create intended sounds, and to evoke associations that are linked to those sounds. For instance, the accumulation of the rhotic /r/ in Goethe's *Erlkönig*, the explosive /t/ in Poe's *The Bells*, or the velar fricative /x/ in Ferdowsi's *Shahnameh*, are excellent examples of how the sound aspect of words can evoke a feeling of toughness, terror, and harshness. And Shakespeare, in spite of Juliet's assertion, also consistently aligns the use of sonorants (i.e., phonemes that are produced with continuous airflow) and obstruents (i.e., phonemes that are formed by obstructing airflow) in his *Sonnets,* with relational meanings defined by the dyad of freedom and constraint, respectively (Masson, 1954; Shapiro, 1998).

The idea that individual vocal sounds or phonemes have standalone meaning is not new. Beyond poetry, archetypal meanings have been associated with sounds, letters of the alphabet, and 'meaningless' strings of characters in many traditions, from the Viking Runes and the Hebrew Kabbalah to the Arab Abjad. Across a variety of religions and belief systems, such as Zoroastrianism, Buddhism, or Taoism, the use of a numinous sound, a syllable, or a group of phonemes, usually known as *mantra,* has been believed by practitioners to have psychological and spiritual powers, symbolizing human longing for truth, immortality, reality, love, peace, and so on. Similarly, associations are made between meaning and single sounds in the ancient Sanskrit texts of Upanishads: "The mute consonants represent fire, the sibilants air, the vowels the sun… The mute consonants represent the eye, the sibilants the ear, the vowels the mind" (Muller, 1879).

In Western philosophy, one of the earliest works addressing the question of the relationship between sound and meaning is Plato's philosophical discussion of the correctness of names in *Cratylus* (Plato, 1892) in which Cratylus explains that an object and its name are naturally connected, whereas Hermogenes expresses the opposite position, saying the only connection between an object and its name is formed by communal agreement. The wise Socrates synthesizes the opposing theories concluding that at the deepest level, names are naturally connected to their referents, but on the surface level convention dictates names.

Despite Socrates' conclusion that arbitrary and non-arbitrary links between sound and meaning co-exist in natural language, the notion of non-arbitrariness has faded from the spotlight of scientific investigations in the past few centuries, and the absolute arbitrariness has been considered a fundamental and necessary design feature of human language (Saussure, 1916/2011; Hockett, 1977). Indeed, the decoupling of the sound structure of words from characteristics of the related referents is assumed to confer a referential power to language, allowing for an infinite number of flexible sound-to-meaning pairings (Gasser, 2004; Hockett, 1958; Monaghan, Christiansen, & Fitneva, 2011). Despite this dominance of arbitrariness in language research throughout the last century, many influential linguists have supported the position of possible synchronic and productive effects of a word's sound on its meaning. For instance, Jakobson (1937) proposed that, "the intimacy of connection between the sounds and the meaning of a word gives rise to the desire of speakers to add an internal relation to the external relation, resemblance to contiguity, to complement the signified by a rudimentary image". That is, the effect of sound in the mind completes its meaning and this, according to Jespersen (1922), may lead to a kind of "natural selection" that "makes some words more fit to survive".

Since the 1920s, there has been a growing amount of empirical work to test the functioning of sound-meaning relationship (usually termed "sound symbolism") in languages. Sapir (1929), for example, raised the issue whether phonemes in isolation are symbolic of differing size by using two nonsense words MAL and MIL. Subjects consistently judged MIL to denote a small object and MAL to a large object. In this vein, the seminal study of Köhler (1947), which demonstrated a consistent link between BOUBA and curvy objects and KIKI and spiky ones, provided empirical evidence of how phonemes alone can convey meaning.

More recently, research from across the cognitive sciences has revealed important patterns of non-arbitrariness in vocabulary, and investigated mechanisms underlying this phenomenon, and brought about an upheaval in how we think about the arbitrary nature of linguistic signs. A growing body of research now challenges the idea of arbitrariness, providing evidence for non-arbitrary sound-to-meaning correspondences by, for example, identifying universal patterns across the more than 6000 languages spoken in the world (Blasi, Wichmann, Hammarström, Stadler, & Christiansen, 2016). These results assign a supplementary function to sound-to-meaning correspondences that structure vocabulary (Perniss & Vigliocco, 2014; Vigliocco & Kita, 2006) and play an important role for both phylogenetic language evolution (Imai & Kita, 2014; Perniss & Vigliocco, 2014; Roberts, Lewandowski, & Galantucci, 2015) and ontogenetic language development (Imai & Kita, 2014; Monaghan, Shillcock, Christiansen, & Kirby, 2014).

Despite the increasing number of studies examining sound-to-meaning associations, a number of questions regarding cognitive and neural mechanisms of this phenomenon remain unanswered. These are the focal points of investigation in the present work. Most importantly, the role of affect as a most basic human experience which shapes the learning, representation, and processing of language (Barrett, 2006; Jacobs, Hofmann, & Kinder, 2016; Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011; Kousta, Vinson, & Vigliocco, 2009; Vigliocco, Meteyard, Andrews, & Kousta, 2009) has been surprisingly neglected. In this work, I will focus on the affective domain, and in particular valence and arousal, which are essential for making basic, critical distinctions between different concepts (cf. Barrett, 2006; Russell, 2003); as empirically established by semantic differential (Osgood, 1952). Such an approach also enables the investigation of the most ancient record of human literature, namely poetry, as the richest source of interplay between sound, meaning, and emotion (Schrott & Jacobs, 2011).

Also, the potential causes of sound-to-meaning correspondences is a central question of this work. That is, whether and in which cases the *motivation* for sound-meaning mapping is based on some inherent qualities of sound, and is derived from perceptual similarities between sound and meaning, i.e., so-called *iconic* relations (see Chapter 2 & 4 for detailed discussions). This includes a measurement of affectivity of the sound of words and texts, and an investigation of how specific properties of the sound play a part in the process of meaning making.

I will focus on the potential relationships between the sound (implicit or explicit) and meaning of words in the affective domain, termed *affective iconicity*. Specifically, I aim to answer the following three main questions. Does the sound of words:

1. evoke affective responses as evident at the behavioral and neural level?
2. influence the processes of affective meaning making and semantic decisions?
3. contribute to the global affective meaning of poems as perceived by readers?

In order to understand the cognitive and neural mechanisms and provide comprehensive answers, I used a variety of approaches consisting of behavioral and neuroimaging experiments, computational modeling, and corpus analysis, combined with different tasks such as rating and semantic decision. Also, I investigated this phenomenon at both the level of single words and the entire text, and in both auditory (spoken words) and visual (printed words) domains[1].

The next four chapters (Chapters 2-5) provide the theoretical background of the work. They emphasize how theoretical and empirical approaches are interconnected and can benefit from a dialogue with each other. The questions raised by the research concerned with iconicity are, on the one hand, rooted in the linguistic, semiotic, and philosophical traditions that address the basic question of how meaning emerges from signs, and, on the other hand, related to emotional, cognitive, and aesthetic processes that can be best understood in the light of empirical research. Thus, substantial progress in answering these questions can only be achieved by means of interdisciplinary approaches.

[1] Note that the term *sound* in this work is equally used to refer to both the sound of spoken words and the implicit sound derived from phonological recoding of visually presented words (Braun et al., 2009; Breen, 2014; Ziegler & Jacobs, 1995). I will discuss this in more details in Chapter 3.

Chapter 2 reviews the most influential linguistic positions on the role of *motivation* regarding the relationship between sound and meaning. The relation to emotion and affect in the respective models will also be discussed. In chapter 3, I will review the important insights and findings regarding cognitive processing models of reading from empirical investigations in neurocognitive psychology in the last decades. As a result, I will then introduce a model of literary reading that links philosophical and linguistic perspectives with empirical results of literature from behavioral and neuroimaging studies, called Neurocognitive Poetics Model (in short NCPM; Jacobs, 2011, 2015a, 2015b). Some of the predictions made by the model will then be used to formulate specific hypotheses for the current investigation. A comprehensive review about the previous research on iconicity will be given in Chapter 4. The limitations identified here will then be used to formulate six hypotheses and research questions that I will summarize in Chapter 5.

The corresponding six studies that empirically investigate these hypotheses and research questions are presented in the empirical part of the thesis (Chapters 6-11). Study 1 (Chapter 6) attempts to address the longstanding conjecture whether phonemes have any inherent affective content or any contribution to words' affective meaning beyond the conventional links in the language system. Study 2 (Chapter 7) pertains to the neural networks underlying the sound of words and investigates whether it can evoke affective brain responses. Study 3 (Chapter 8) extends the scope of findings from Study 1, and asks whether a congruence between sound and meaning (iconicity) can facilitate evaluative decisions on words' affective content. Study 4 (Chapter 9) focuses on the neural mechanisms of affective iconicity and aims to answer how the affective information of words from two different sources (i.e., sound and meaning) are integrated in the human brain. In Study 5 (Chapter 10), and Study 6 (Chapter 11) the effect of iconicity is investigated at the text level. Study 5 provides the methods and tools for extracting relevant phonological units in texts. This approach will then be used in Study 6 to examine whether and to which extent the overall affective meaning of a text is (co-)determined by the specific use of particular phonological units.

Finally, Chapter 12 will give a comprehensive summary of the results of the six empirical studies. Conclusions will be made on how the results improve our understanding of the role of sound in both general processes of meaning making, as well as in affective and aesthetic processes of literary reading. Based on the findings of this dissertation, extensions and upgrades of existing models and hypotheses will be suggested, and the limitations of the current investigation as well as the implication for the future research will be discussed.

Chapter 2

Linguistic Perspectives

How meaning emerges from words seems at once the most obvious and the most obscure question about words. It is obvious because it is what we use language for, i.e., to communicate with each other, to convey 'what we mean' effectively. Nevertheless, this question has been considered in controversial and multifaceted ways. Although a number of prominent thinkers and philosophers over the last hundred years introduced and advanced excellent theories addressing this question—from Frege, Bréal, and Husserl, to Cassirer, and Wittgenstein—none of these theories has been able to remove all the difficulties inherent to the study of semantics in human language. Focusing on some of the most noteworthy attempts relevant to the empirical investigation conducted in this dissertation project, I will give a brief overview of these theories to shed light on the original question of *how meaning emerges from words*, and, more specifically, on the question of *motivation*. Assuming a simple model consisting of a word and an object, I shall use the term *motivation* to describe the extent to which the object determines the word (cf. Fischer & Nänny, 2001). The more a word is constrained by the concept, the more motivated the word will be. Therefore, in the context of sound-meaning relationship, the question I aim to answer can be rephrased as: *Are words motivated, and if so, to what extent?*

2.1 Words as Motivated Signs

To give a more comprehensive picture of the possible answers to the aforementioned question, I would like to emphasize the understanding in which words are primarily considered as *linguistic signs*, or even more generally as *signs*. This way of looking at words will enable us to include valuable insights from the study of signs which has been conducted throughout the history of both philosophy and psychology.

The scientific study of signs began with American philosopher Charles Sanders Peirce (1839-1913), the founder of *semiotics* (the term was coined by him to refer to the study of signs; Peirce, 1931). Peirce was interested in how we make sense of the whole world around us, and thus was less concerned with the linguistic aspect of semiotics. Other than Peirce, there were two key figures in the early development of the study of linguistic signs. The first one was the well-known Swiss linguist Ferdinand de Saussure (1857-1913); the founder of modern linguistics and of what is now more usually referred to as *semiology* (first used in his Course in General Linguistics; Saussure, 1916/2011). The second pioneer was the German psychologist, philosopher and linguist Karl Bühler (1879-1963), who provided a pioneering theory of *semiosis* (Bühler, 1934) and presented a multidisciplinary approach for understanding language by an excellent synthesis of methods and insights from empirical psychology, philosophy, linguistics and communication sciences.

In the remaining part of Chapter 2, I will start with the linguistic model of Saussure, and will compare his model with the one proposed by Peirce in his semiotics theory. Next, I will juxtapose the so-called Organon model of Bühler with the models of Saussure and Peirce. In doing so, I aim to address a central question of my investigation into the sound-meaning relationship; namely the degree of *motivation* in linguistic signs. In a subsequent step, I will introduce a synthesis of these traditions as proposed and elaborated in the theories and the works of one of the most influential linguists of the century, Roman Jakobson (1896-1982), which is highly relevant to this dissertation project. Jakobson was the first to develop a model that describes the specific functions of language and literary forms of literature and poetry, and his investigations into sound-meaning relationship provide a theoretical framework that continues to inform current language models (e.g., the so-called Neurocognitive Poetic Model, see Chapter 3).

2.2 Saussure vs. Peirce, or Semiology vs. Semiotics

The principal concept of Saussure's theory is a dichotomy in which a sign consists of two focal components; the *signifier,* or what he defines as the sound pattern; and the *signified,* or the concept (Saussure, 1916). The signifier refers to something that is in a material form; it is physical, explicitly exists, and can be distinguished by human senses. On the other hand, the signified denotes something that literally and physically does not exist (Eco, 1976). Saussure asserted that both the signifier and the signified are closely related and complement each other. That is, both cannot be separated, nor can one aspect exist in the language system without the existence of the other (Figure 2.1, top level).

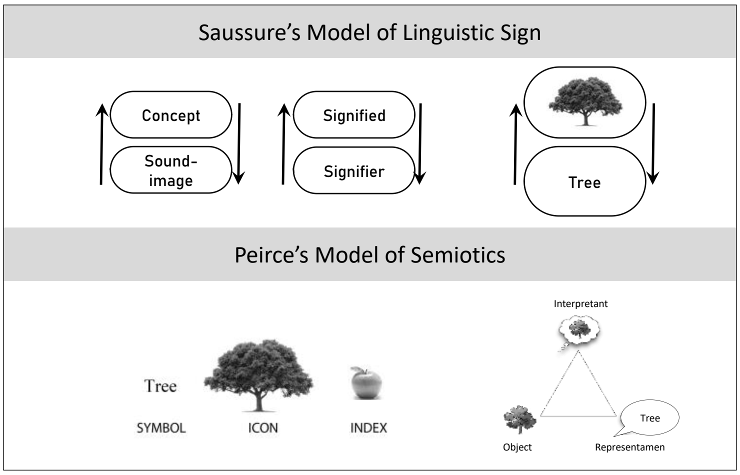


Figure 2.1. A comparison of two linguistic models as exemplified by a 'tree'. Top: the dichotomous model of Saussure, Bottom: the trichotomous model of Peirce

In contrast to the dichotomous concept of Saussure's theory, Peirce's theory of signs focuses on a three-dimensional system. This consists of i) the *sign* (representatum), ii) the *object,* which is also referred to as referent, and iii) the *interpretant,* who interprets the relationship between sign and object. The first aspect in this model is synonymous with Saussure's concept of the signifier, which refers to a physical sign (e.g., a word). Saussure's concept of the signified, however, is divided by Peirce into two components, namely object and interpretant. According to this model, meaning does not emerge through a relationship between a sign and an object, as assumed by Saussure, but through the way we perceive or understand a sign and its relationship to the object that it is referring to.

Besides this difference, there is a more important divergence between these two models which is highly relevant to the present work, and to the question of motivation. In Saussure's semiological perspective, the only relationship that links the signifier to the signified is a mutual agreement between the sender and the receiver of the sign. Accordingly, any sign is subject to the conventional system and there is no essential or natural reason why a particular signifier should be attached to a particular signified (see *l'arbitraire du signe* - the arbitrariness of the sign - by Saussure). A direct consequence of this proposal is that the sound of a word per se does not play any contributing role in its meaning. Thus, a linguistic sign, according to the Saussure model, is completely *unmotivated*.

Unlike Saussure, Peirce proposed that signs could be assigned to three definitional categories (Figure 2.1). As an *icon*, the sign (representamen) resembles or imitates its denoted object in that it possesses some of its qualities. Therefore, the relationship between, for example, how the sign sounds and what the sign stands for—its referent and the sense behind it—is marked by similarity. For example, a picture of a tree, a portrait, a sound effect, a cartoon, or a statue, are all examples of iconic signs referring to a real tree, a real person, a real sound, and so on. As an *index*, the relationship between the sign and what it stands for may have to be learned. The link between the representamen and its object may only be inferred via causality. For instance, smoke being an index for fire, thunder for rain, a door bell for ringing, etc. And finally, a *symbol* is assigned arbitrarily to its object, or is accepted due to common convention. Therefore, the relationship between the representamen and what it stands for, in the symbolic case, must be learned. Letters of the alphabet, the number system, or mathematical signs are examples of symbolic signs.

In contrast to the Saussure model, this model distinguishes between iconic signs which are highly motivated, and symbolic signs which are completely unmotivated. An iconic relationship based on similarities between the signifier and the signified is acknowledged only in the Peirce model. However, it's important to note that Peirce's interpretation of semiotics as a study of signs extended far beyond linguistic signs. Indeed, he conceived semiotics as a synonym for the concept of logic. Therefore, the main constraints of Peirce's theory are the human mind and sign boundaries, and it was Jakobson (1965) who first discovered the relevance of iconicity to linguistics. He drew the attention of linguists to three categories of iconicity, which I will discuss later in this chapter.

There are some limitations that both of the above models share. For instance, both models do not explain how abstract signs or those with grammatical meanings (i.e., function words) represent an object (Lehmann, 2007). Moreover, as language is primarily considered a cognitive and logical undertaking in both of these models, the role of emotion, and emotional communication is largely neglected. If someone, for instance, cries in pain, the sign, in Peirce's terminology, would be an index; but what is the object it represents? Likewise, in the Saussure's model, it is stressed that the signifier (e.g., sound pattern) and the signified (concept) are the components of the sign that are 'intimately linked' in the mind 'by an associative link' (Saussure, 1916). For Saussure, language has no origin in a domain of preexisting ideas or concepts. Signs can exist only in opposition to and in relationship with other signs, and through this they earn value. He emphasizes that there is no 'inner relationship' between the signifier and the signified; rather this relationship is *unmotivated* and arbitrary (Saussure, 1916, p.69). According to this notion, the content of the sign originates from a relation between the respective signs within a system. Therefore, all signs obtain their meaning in relation to other signs. However, this is problematic for the question of the signs' origin, as every sign you begin with would only lead to another sign, and that sign to yet another; an infinite, endless chain (c.f. Marlin, 2008).

These limitations clearly show that theories and models which consider language as a static self-contained system neglect a significant point essential for emerging meaning: i.e., the psychological and biological characteristics of language and its ultimate goal of serving as a communication system. This is where *semiosis* comes into play. Semiosis, much more than semiotics and semiology, emphasizes the communicative role of language. It defines every linguistic act as consisting of a *sender* willing to influence a *receiver* by conveying a perceptible *signal*. Importantly, the desired influence on the receiver does not only comprise mental attitudes—as recognized by Saussure and Peirce—but also emotional attitudes, as well as actions and behaviors in the broadest sense (cf. Kirsner, 1985). One of the most comprehensive models that emphasizes the communication function of language in the framework of semiosis and also provides interesting insights into the degree of motivation at different levels of communication is the so-called Organon model proposed by Karl Bühler (1934).

2.3 Bühler's Organon Model

Karl Bühler's contribution to the early scientific study of language is not very well known, certainly not in the English-speaking world. This is chiefly because his master work *Sprachtheorie* (Bühler, 1934) was not published in English until 1990, although the first German edition appeared in 1934.

By emphasizing the communicative function of language, Bühler developed a model in which language is not reduced to a system of signs from which *meaning* simply emerges. Rather, the semantics of human language, according to Bühler, is rooted in social interaction and practical human action. In his theory, Bühler describes how words can evoke sensory, motor, and affective memories related to their meaning. In doing so, they activate the same experiences as the corresponding natural events. This pioneering way of looking at language—together with the ingenious ideas of Ludwig Wittgenstein—is perhaps the origin of "Embodied Cognition": a notion that has been of enormous influence to existing theories and paradigms on the representation of semantic knowledge (e.g., Aziz-Zadeh & Damasio, 2008; Barsalou, 2008; Gallese & Lakoff, 2005; Glenberg & Kaschak, 2002; Harnad, 1990; Jacobs, Hofmann, et al., 2016; Zwaan, 2004; see Jacobs, 2015a, for more details on this view).

This embodied view on the emergence of semantic meaning in human language enables us to broaden our understanding of how linguistic signs can be motivated. Both the relationship between the sign and the object and the physical properties that the sign possesses (e.g., the sound) can be linked to the sensory, motoric, or affective experiences and memories that form the overall meaning of the word. Thus, a greater embodiment in language processing would take place when the word form can directly trigger aspects of modality specific experience. This is an important characteristic of Bühler's model that was left unaddressed in the models proposed by Saussure and Peirce.

In his Organon model (Figure 2.2), Bühler distinguishes three elements that are necessary for communication to occur: i) a *speaker* or *sender*, ii) a *hearer* or *receiver*, and iii) *objects* or *states of affairs* that are linguistically referred to (*the designatum*).

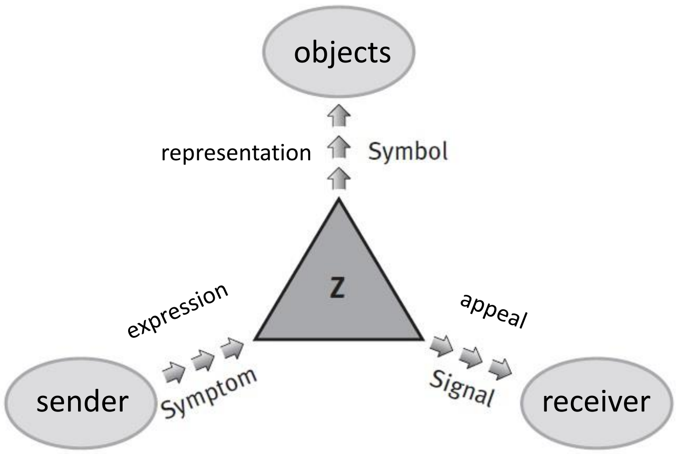


Figure 2.2. Organon model of language as proposed by Karl Bühler (taken from Schrott & Jacobs, 2011)

In this model, Bühler suggests that language is mainly used to fulfill three major communication functions, and that these communication functions, in turn, can explain linguistic functions: i) *Ausdrucksfunktion* or the expressive function, in which the sign expresses an aspect of the sender; in this capacity it is a *symptom*, ii) *Appelfunktion*, or the appealing function, in which the sign appeals to the receiver; in this capacity it is a *signal*, and iii) *Darstellungsfunktion*, or the representative function, in which the sign represents a designatum; in this capacity it is a *symbol*.

An additional difference of Bühler's model to those provided by Saussure and Peirce is in the notion of designatum, which stands for the content of a communication. Unlike Saussure's and Peirce's models, the content of a message, as described in the Organon model, is not solely in the sign per se. Rather, it is the sum of its expressive, representative and appellative functions. For instance, when a sign expresses the (inner) state of the sender, it does so because the sender produces the sign, and consequently has a causal relation to the sign. In this sense the sign is *motivated* by the sender. With regard to the main question of this dissertation, this means that the affective state of the sender is in any case reflected in designatum. From a modern psychological perspective, affective states of the sender are linked to specific physiological states which will be reflected in the vocal behavior of the sender and thus extend to acoustic features of the speech signal in the form of, for example, implicit or explicit affective prosody. On the other hand, a sign can potentially represent the designatum by resembling specific aspects of it. This represents a clear case of motivation recognizable from the Piercean understanding of iconicity. And lastly, a sign can appeal to the receiver through the conventions held between sender and receiver. The receiver may be able to empathize with the sender and thus interpret the sign on the basis of its symptomatic relation to the sender (see Lehmann, 2007, for a detailed discussion on this view). In the latter case the sign is motivated by the receiver on the basis of mutual conventions and agreements. This way of looking at the linguistic sign—in which different functions and layers shape the linguistic meaning in an interactive way—goes beyond the simple dichotomous or trichotomous models proposed by Saussure or Peirce. It can explain, for instance, why even highly iconic signs, such as onomatopoetic expressions for the same concept, can largely differ across different languages, e.g., the words for the noises that pigs make vary from "buubuu" in Japanese to "rok-rok" in Croatian (Farmer, Christiansen, &

Monaghan, 2006): Hence, the representative function (i.e., the similarity between word form and concept) and the appealing function of language (social agreements in the form of phonological constraints in the respective language) are combined with each other and thus help the receiver to encode the meaning of proposed onomatopoetic expressions.

Obviously, the three sign functions in Bühler's model share some similarities with Peirce's three categories of signs in a way that can be represented by mapping Bühler's sign functions onto Peirce's classes of signs as shown in Table 2.1. Importantly, unlike Peirce's model, which considers motivation in language only for iconic signs, signs in Büler's categorization can be motivated at all three types of relation, thus providing a comprehensive model for the level of *motivation* in linguistic sign.

Saussure	motivated
Signifier	—
Signified	—

Peirce	motivated		Bühler	motivated
Icon	✓	≈	Symbol	✓
Index	—	≈	Symptom	✓
Symbol	—	≈	Signal	✓

Table 2.1. The degree of motivation of linguistic sign in light of three different language models (i.e., Saussur's, Pierce's, and Bühler's model)

2.4 Jakobson's Quest for the Essence of Language

Last but not least, I will introduce the model of language of Roman Jakobson, who has generated a brilliant synthesis of the above mentioned models and theories throughout a broad range of works devoted to the study of language and literature. Jakobson was strongly influenced by Saussure's ideas and was a key figure in Russian structuralism. This was evident in his collaborative work with Lévi-Strauss in the 1960's (Lévi-Strauss, 1960) and in his contribution to Russian formalism (e.g., Shklovsky, 1990; Trotsky, 1957), which mainly focused on the distinction between the poetic and prosaic use of language. Jakobson was also inspired by the works of Peirce, who he discovered after he went to the United States, appraising Peirce as "the most universal and inventive of American thinkers" (Jakobson, 1965). Finally, for his work on the functions of language, Jakobson was so strongly influenced by Bühler's Organon model that his own model is usually recognized as an extension to Bühler's model.

As one of the important founders of Russian formalism, Jakobson was deeply interested in poetry and literature and looked intensively at the structures and devices that literature employs. In an attempt to provide a model for literary reading, Jakobson presented an extended version of the Organon model to which he added the so-called poetic, phatic and metalingual functions. In the poetic function, communication focuses on "*the message for its own sake*" (Jakobson, 1960b). It is the dominant function in poetry and any aestheticized form of language. In the phatic function, communication focuses on the *act of contact*, e.g., expressions such as "Hello?", "OK?", etc. Finally, if the communication is oriented towards the *code of communication*, it is metalinguistic, e.g., "do you understand me?"

Besides the development of this model of language, and in particular, the definition of the poetic function, Jakobson applied Peirce's concept of iconicity to literary studies and poetry and emphasized the role of sound-meaning relationship as a general feature of language. In the quest for the *differentia specifica* of poetic texts, Jakobson points to iconic properties as a promising direction for research (Jakobson, 1965). In doing so, Jakobson proposed that the distinction between form and meaning drawn by structuralists (e.g., Saussure) and generativists (e.g., Chomsky) were not entirely valid. In his most important work on this subject 'Quest for the Essence of Language' (Jakobson, 1965), Jakobson merged Peirce's semiotics into linguistic theory, and used Peirce's classification of signs and associated terminology. Here, he distinguished three different types of iconicity, i.e., imagic, digarammatic, and metaphoric, and provided instances for each of them in syntax, phonology and morphology. By exemplifying Caesar's famous "*veni, vidi, vici*", he identified a new level of iconicity at the textual level to show how "*the temporal order of speech events can tend to mirror the order of narrated events in time or in rank"*. His approach, and in particular his emphasis on the phonological level of language in poetry, created an excellent basis for an area of current empirical study known as (neuro)cognitive poetics (Jacobs, 2011, 2015a, 2015b; see Chapter 4 for details).

To sum up, the idea of iconic signs from Peirce helped Jakobson resolve the paradox between Saussure's arbitrariness on the one hand and the very general existence of sound-meaning relations on the other. Therefore, Jakobson's model of language is both the most comprehensive model and the most relevant to my investigation in the present dissertation: i.e., it takes into account i) the role of iconicity as a general feature in language, ii) the role of emotion in language and human communication, and iii) the evaluation and interpretation of literature and poetry. A simple matrix of how each theory accounts for these three topics is summarized in Table 2.2. Obviously, despite the insightful characteristics of the introduced models, none of them is formulated in a way that could make predictions that are empirically testable, e.g., predictions about cognitive, psychological, and neurophysiological aspects of language processing.

Model	Accounting for			
	Iconicity	Emotion	Literature	Empiricism
Saussure	–	–	–	–
Pierce	✓	–	–	–
Bühler	✓	✓	–	–
Jakobson	✓	✓	✓	–

Table 2.2. A summary of the proposed models and theories of language, and their relation to the topics relevant to the present work.

2.5 Conclusion

Despite the limitations of the models presented and theories provided, all significantly enhanced our understanding of the factors and mechanisms that lead us to make sense of linguistic signs. Therefore, the initial question of *how meaning emerges from words* can be partly answered by considering the underlying substitutes of linguistic signs, the role of sender, receiver, and the contextual factors during communication. Also, the role of formal features of a sign (e.g., sound of a word) and their standalone contribution to the meaning of a linguistic message (e.g., word, text) can be understood more readily in the light of the different language models proposed above.

Theoretical investigations into how human language works have continued to garner attention and build upon the theories and models of the aforementioned scholars. The most important 20th century thinkers—among them Bertrand Russell, Ludwig Wittgenstein and Martin Heidegger—shifted their focus away from *ideas* in the mind to the *language* in which thinking is expressed. The initial ideas and models proposed by structuralists such as Saussure, Jakobson, and Levi-Strauss were strongly criticized or rejected, leading to a new era of *post-structural* ideas and attitudes with a range of new claims and statements about the nature of human language, e.g., Roland Barthes' the death of the author (Barthes, 1994), Michel Foucault's the structure of power (Foucault, 1982), or Jacques Derrida's Deconstruction (Derrida, 1976), just to name a few examples. Although the obsession with language in the era of postmodernism contributed significantly to its study, it made linguistic theories in many cases only a matter of philosophical debates which relied on reason as the chief source of knowledge. Thus, the role of empiricism as a fundamental part of the scientific methodology in the formation of human knowledge faded more from the spotlight of philosophical theories.

As a researcher with a strong focus on the empirical investigation of the topic, I will therefore take all the insights, knowledge and inspiration accumulated through proposed models and theories and, in the following chapters, proceed to discuss empirical evidence from related scientific disciplines. Recent advances in technology and experimental techniques have made it easier to gather the necessary data for empirical approaches to understanding human language. Therefore, a synthesis of a theoretically informed and empirically based approach will enable us to benefit from the intellectual hypothesis of the theoreticians as well as the empiricist's ability to deliver factual answers.

Since the ultimate goal of this dissertation is to provide a better understanding of processes underlying literary reading I will focus on empirical evidence pertaining to different processing levels necessary for *reading* a text in the next chapter. I will start with the most basic level of single word recognition and move toward the consideration of the complex affective and esthetic processes which underlie the comprehension of an entire text.

Chapter 3

Reading: from a Single Word to a Whole Text

When reading a text, several complex processes such as orthographical, phonological, morphological, semantic and syntactical information processing, global text comprehension and affective-aesthetic processes, need to be integrated into a whole in order to have a successful reading experience (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Jacobs & Hofmann, 2011; Perry, Ziegler, & Zorzi, 2007; Price, 2012). A number of psychological models in the last decades attempted to describe how this rather recent cultural technique (Coltheart et al., 2001; Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981; Seidenberg & McClelland, 1989) can be acquired by children and mastered in the course of development (e.g., Frith, 1985, 1986). In order to have a better understanding of this multifaceted phenomenon, previous research aimed to reduce the complexity of reading research by breaking it up into its sub-components, starting with the question of how a single word is recognized (Grainger & Jacobs, 1996; Jacobs & Ziegler, 2015). Therefore, I will first present the current status of research regarding single word recognition at both the level of psychological processes and the level of the neural correlates and substrates. I shall then proceed to address the subsequent processes of the integration of words into sentences and the whole text.

3.1 Single Word Processing

Visual word recognition is typically divided into orthographic, phonological and lexico-semantic processes (Fröhlich et al., 2018; Grainger & Jacobs, 1996; Seidenberg & McClelland, 1989; Ziegler et al., 2008). At the orthographical level, letters are recognized and integrated into larger sublexical units. The next two steps in the process of word recognition, namely pronouncing the word and accessing the lexical meaning, are typically understood through widely accepted dual-route models (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart et al., 2001). These models were first conceived by Forster & Chambers (1973) and advanced in the same year by Marshall & Newcombe (1973). This theory suggests that two separate mechanisms are involved in reading. One mechanism, termed the (fast) lexical route, is the process used for regular words. Via this route, skilled readers can recognize known words through the direct activation of the word in a hypothetical *mental lexicon*. The nonlexical or sublexical route, on the other hand, is the process that can be thought of as a letter-sound rule system that allows the reader to actively build a phonological representation and read the word aloud. This route, according to the model, is used for the words that are not stored in the lexicon.

Several computational models of visual word recognition, such as the dual process model (COP+; Perry et al., 2007) and the multiple read-out model including phonology (MROM-p; Jacobs, Rey, Ziegler, & Grainger, 1998) rely on the idea that semantic access depends on phonological activation. Thus phonology would be typically computed before people access the meaning of a word (see also Braun, Hutzler, Ziegler, Dambacher, & Jacobs, 2009). Indeed, research on visual word recognition in the last two decades has provided behavioral, computational, and neuroimaging evidence that phonological information is automatically generated from the printed word during silent reading, providing an early and major constraint for lexical access (e.g., Braun et al., 2009; Coltheart et al., 2001; Conrad, Carreiras, Tamm, & Jacobs, 2009; Conrad, Tamm, Carreiras, & Jacobs, 2010; Hofmann, Stenneken, Conrad, & Jacobs, 2007; Ziegler, Ferrand, Jacobs, Rey, & Grainger, 2000; Ziegler & Jacobs, 1995). In line with this idea, it has been shown that letter–speech sound associations do not develop in parallel with visual letter recognition, but work in concert to form orthographic–phonological bonds which remain active even in experienced reading (Blomert, 2011).

The first connectionist model to really focus on the dynamics of information processing in the different mentioned layers was the interactive activation and competition model (IAM) proposed by McClelland & Rumelhart (1981). In this model, perception results from excitatory and inhibitory interactions of the so called *detectors* for visual features, letters and words. The model made all information processing steps between input and output fully transparent, and achieved a gold standard of model evaluation criteria (Jacobs & Grainger, 1994). This model inspired a number of follow-ups based on a similar architecture. These included the Dual Read-Out Model (DROM; Jacobs & Grainger, 1994), and the Multiple Read-Out Model (MROM; Grainger and Jacobs, 1996), which provided novel explanations for the lexical access of printed words (see Hofmann & Jacobs, 2014, for a review).

At the cerebral level, processes of visual feature extraction and orthography take place in the occipital pole (OP), and in an area deep in the occipitotemporal cortex known as the visual word form area (VWFA; Warrington & Shallice, 1980, see Hannagan, Amedi, Cohen, Dehaene-Lambertz, & Dehaene, 2015; Wandell, 2011, for recent reviews). Further processes take place in the supramarginal gyrus (SMG), which is associated with the phonological representation of printed words (Carreiras, Armstrong, Perea, & Frost, 2014; Price, 2012). For the lexico-semantic aspect, the inferior frontal gyrus (IFG) has been shown to be involved in the integration of high-level linguistic properties such as semantics, and in providing feedback connections to constrain the orthographical representations of the fusiform area (Binder, Desai, Graves, & Conant, 2009a; Carreiras et al., 2014; Woodhead et al., 2014).

Figure 3.1 presents a simple integration of insights from interactive activation models of visual word recognition and neuroimaging data (see Carreiras et al., 2014, for a more comprehensive integrative model). As current research in this field suggests, the only connection between phonology and semantics is based on associative links in memory, which are determined by the language system.

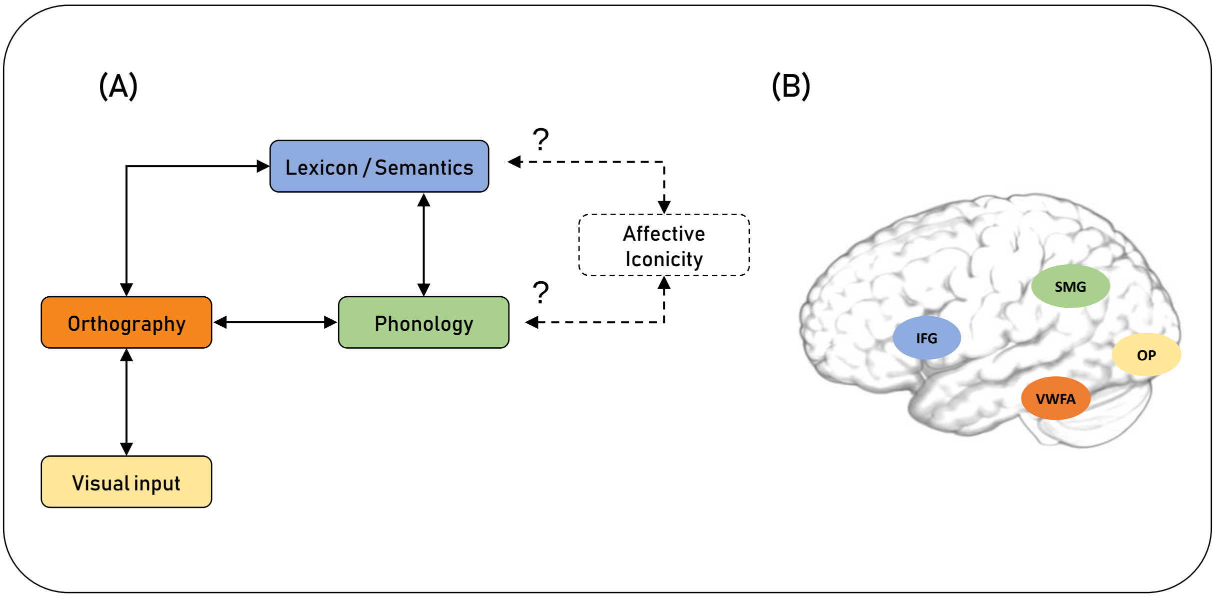


Figure 3.1. A) A simple presentation of an interactive activation model of visual word recognition. The notion of affective iconicity suggests an additional route linking affective aspects of phonology to affective aspects of semantics beyond the associative links determined by the language system. **B)** The most important brain regions associated with different processing levels: OP, occipital pole; VWFA, visual word form area; SMG, supramarginal gyrus; IFG, inferior frontal gyrus.

However, the hypothesis of affective iconicity suggests that the phonology of words has per se affective iconic properties which can interact with affective aspects of semantics. Thus, in this dissertation, I attempt to show that the assumed single connection between phonological and semantic representations of words needs to be revised. More specifically, I aim to answer the questions of whether, how, and in which brain areas the effect of words' sound emerging from the recoding of phonological information interacts with aspects of words' meaning. I shall integrate the findings and the results of my investigations in a general reading model (as displayed in Figure 3.1) and introduce an updated model that accounts for the phenomenon of affective iconicity.

3.2 Sentence Processing

Moving from the processing level of single words to the level of the whole sentence, psychological models of reading have to predict a level of language comprehension far beyond that of isolated, single words. An appropriate model operating at this level should be able to explain the variations in sentence comprehension based upon a number of varying elements: from syntactic processes (e.g., grammatical rules) and contextual variables (e.g., knowledge about the writer) to reader-related issues (e.g., reader's background knowledge). Early models of syntactical processing relied on the fact that during the reading process the reader makes predictions about the upcoming words. This idea inspired a range of models focusing on a so-called N400 component, which reflects semantic competition during sentence processing (Kutas & Hillyard, 1984). This component is defined by cloze completion probabilities based on an observed correlation between the amount of typical completions of a sentence fragment and the size of the N400 (Dambacher, Kliegl, Hofmann, & Jacobs, 2006; Kutas & Hillyard, 1984). Event-related potential studies have also revealed the temporal resolution of this effect by showing that the comparison of expected and incoming words during reading is an extremely rapid process: it takes no longer than 90 ms after visual input (Dambacher, Rolfs, Göllner, Kliegl, & Jacobs, 2009). Also, eye-tracking data suggest that readers routinely use a variety of sources of probabilistic information—from phonological cues to syntactic context and real-world knowledge—to anticipate the processing of upcoming words (e.g., Staub & Clifton, 2006, see also Christiansen & Chater, 2015, for a "Chunk-and-Pass" view).

3.3 Text Processing

At the text level, i.e., reading a story, a novel, or a poem, the contextual information and the reader-related issues become even more crucial for understanding and remembering a text. According to one of the most influential models of text comprehension, readers integrate their knowledge of the world with information in the text. This results in the information available for readers being much richer than what is provided by the text alone (Kintsch & van Dijk, 1978). Such a view is rooted in the notion that the meaning of a text is not an object to be found within a text—rather it is an event of construction that occurs between the text and the reader (Iser, 1976). In his interactive theory of reading, described in “der Akt des Lesens” (Iser, 1976), Wolfgang Iser attempted to look at processes by which readers interact with texts. In short, the theory states that during reading the reader forms a representation of the text that is continually altered by new information. ‘Situation models’ have been proposed based on a similar idea. These models function by maintaining and updating representations of information that is presented in a story (Kintsch & van Dijk, 1978; Zwaan & Radvansky, 1998). Typically, there are five dimensions that are assumed to constitute the situation model: i.e., time, space, protagonist, causation, and intentionality (Zwaan & Radvansky, 1998). In line with the predictions made by situation models, recent neuroimaging studies have shown that readers dynamically activate specific visual, motor, and conceptual features of activities while reading about analogous changes in activities in the context of a narrative (e.g., Speer, Zacks, & Reynolds, 2007; Yarkoni, Balota, & Yap, 2008).

3.4 The Historical Neglect of Emotion in Cognitive Sciences

Despite the specification of emotional aspects in earlier language models such as Bühler’s Organon (1934) and their extension by Jakobson (1960; see Chapter 2), most of the models discussed focus solely on cognitive processes and fail to address the contribution of affective processes to the act of reading (Jacobs, 2011). Similarly, most emotion theories have not tackled language-related processes (but see Koelsch et al., 2015, for a recent exception), leading to what has been called a *double neglect* (Jacobs 2011, 2015a, 2015b). According to this view, reading cannot be reduced to *cold* information processing (Jacobs, 2015a, 2015b; Jacobs et al., 2015). Rather, it involves affective and aesthetic processes which need to be considered in current models of language at different levels of processing, from word recognition and sentence processing to text comprehension. In order to bridge this gap and to investigate such *hot* information processing, a recent Neurocognitive Poetics Model of literary reading has been proposed (Jacobs, 2015a) which emphasizes the aesthetic and affective function of language in literature reception. Since this dissertation project investigates the role of affective iconicity in both ordinary use of language and literary reading, I will shortly review previous empirical research in a field devoted to the empirical study of literature and poetics (i.e., cognitive poetics). I shall then focus on a recent development in this field and on the aforementioned model of neurocognitive poetics, which attempts to link the findings of cold reading processes, and of classic cognitive poetics, with recent psychological and neurophysiological insights about the emotional and aesthetic processes which underlie literature reception.

3.5 A Neurocognitive Poetic Model

Over the last two decades, the study of literary reading has been complemented by developments and insights into cognitive linguistics, and particularly by research into an area at the interface between linguistics, literary studies and cognitive psychology known as cognitive poetics (e.g., Gavins & Steen, 2003; Miall & Kuiken, 1994; Stockwell, 2002, 2007; Tsur, 1992a), or cognitive stylistics (e.g., Semino & Culpeper, 2002; Semino, 2009). In contrast to theoretical and philosophical approaches from classic linguistic and literary studies, cognitive poetics deals with the central question of how empirical *readers* comprehend and interpret the language of literary texts by conducting *experiments* "with different types of literary discourse, in different reading contexts with different kinds of readers" (Van Dijk, 1979) with the ultimate goal of developing models of text comprehension and production (e.g., Kintsch and van Dijk, 1978).

The developments and insights from this relatively young discipline have been recently complemented by the inclusion of methods and models "for investigating the neurocognitive and affective processes associated with processing and experiencing literary texts", as recently proposed by Jacobs' Neurocognitive Poetics Model' (in short NCPM, (Jacobs, 2015a, 2015b; Jacobs & Hofmann, 2011; Jacobs & Willems, 2017; Nicklas & Jacobs, 2017). The model has been developed to make predictions concerning behavioral, neuronal and cognitive-affective responses of readers based on empirical evidence and results from studies on poetic forms and evaluation, text processing, emotion, and neuroaesthetics. The NCPM is a pioneering model in this field, and not only advances our knowledge about literature reception, but also helps provide a better understanding of different processes in the human brain, such as mental simulation, emotion, empathy, and immersion (Jacobs & Willems, 2017; Willems & Jacobs, 2016).

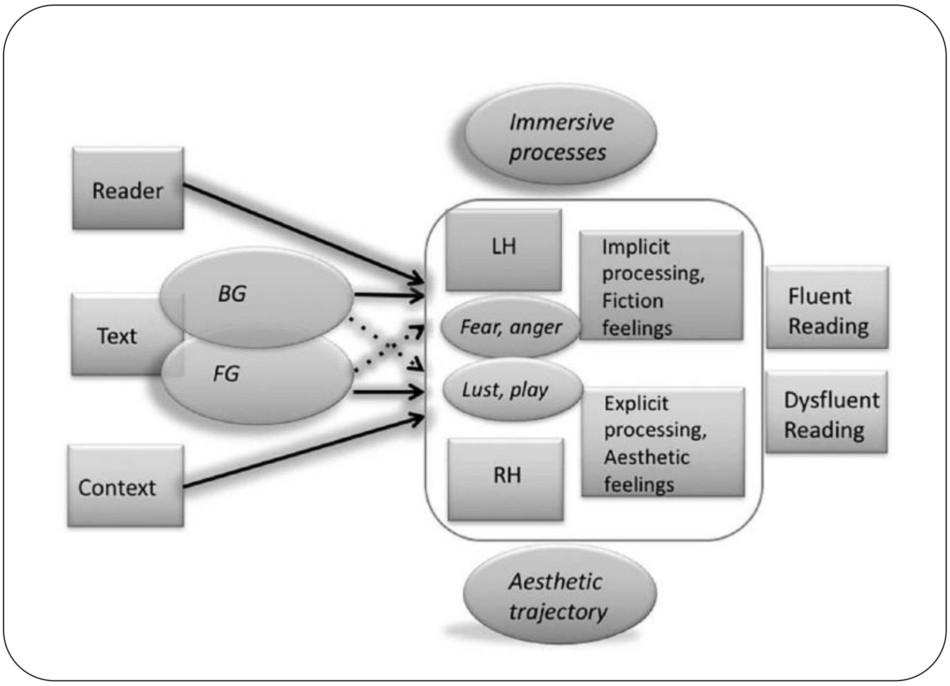


Figure 3.2. The simplified neurocognitive poetics model of literary reading by Jacobs (taken from Jacobs, 2015a).

The NCPM (Figure 3.2) integrates insights from the long tradition of linguistic and literary studies (e.g., Jakobson) with the latest neurobiological findings on emotion and affective processes (e.g., Panksepp). That is, it includes i) a novel hypothesis of *fiction feeling* regarding the phenomenon of immersion during reading (Figure 3.2, upper route) together with ii) the notion of *foregrounding* covering the process of an aesthetic trajectory (Figure 3.2, lower route), as well as iii) a neuroscientific view that bridges the language-emotion gap, termed the *Panksepp-Jakobson-hypothesis*. These points will be briefly introduced in the following sections.

3.5.1 Fiction Feeling Hypothesis

The major part of the reading process takes place in the upper route of the model, i.e., "background" reading. This includes the highly automatized part of the fluent reading, starting from word recognition and sentence processing, up to building a situation model. According to the model, the related processes of background reading are localized in the left hemisphere reading networks including the VWFA, the classical Wernicke and Broca centers, and the prefrontal cortex. In this type of fluent reading, text elements match reader's predictions and anticipations, thus keeping the reader in a mode of so-called immersive reading. A *Fiction Feeling Hypothesis* lies at the core of this route, stating that narratives with emotional as opposed to neutral content encourage readers to be more empathic towards the protagonists and thus engage with the affective empathy network of the brain (Jacobs & Lüdtke, 2017; Jacobs & Schrott, 2015). That is, fear-inducing passages or descriptions of protagonists' pain would cause an increase in the involvement of the core structure of pain and affective empathy in the reader which, in turn, can result in a higher degree of immersion, as empirically shown by Hsu, Conrad, & Jacobs (2014) in a neuroimaging study on Harry Potter passages.

3.5.2 Foregrounding

Foregrounding is essentially used as a stylistic technique—from Shklovsky's (1925/1990) Russian term *ostranenie*—used to defamiliarize the reading experience in textual composition through stylistic distortion of some sort, either through an aspect of the text which deviates from a linguistic norm or, alternatively, where an aspect of the text is brought to the fore through repetition or parallelism (Simpson, 2004, Jacobs 2011, 2015a, 2015b). In general, foregrounding refers to a form of textual patterning and is capable of working at any level of language, whether phonological, morphological, syntactical or semantic. The analogy or correlation between the idea of foregrounding theory in literature and the Gestalt notion of figure-ground (Rubin, 1921; Köhler, 1929; Koffka, 1935) has led to the application of Gestalt principles to literary studies. However, it was only after empirical results demonstrated that literary foregrounded elements are related to more intensive and extensive cognitive processing (van Peer, 1986), increased memory for stylistic features (Zwaan, 1993) and deeper emotional experience (Miall & Kuiken, 1994; Hakemulder, 2004), that the concept of foregrounding in modern text processing was accepted "as a manifestation of the Gestalt psychological principle of figure-ground discrimination" (van Holt & Groeben, 2005).

Recent neuroimaging evidence has provided further support for this notion by showing that foregrounded items in a text can increase cognitive processing demand and activate brain areas related to affective processing, thereby setting the reader into a mode of aesthetic perception (Bohrn, Altmann, & Jacobs, 2012a; Bohrn, Altmann, Lubrich, Menninghaus, & Jacobs, 2012b). These findings provide empirical support for the role of foregrounding in the NCPM. When the reader encounters more foregrounded elements in a text, "mixed feelings, aesthetic emotions, and (self-) reflective thoughts oust the general feeling of familiarity", engaging the lower (slow) route of reading in the model (Jacobs, 2015a). These processes are assumed to correlate with a slow reading mode and significant neural activity in right hemispheric networks. Relevant to the present work, at the phonological level foregrounded elements such as alliteration or rhyme can cause a conscious sub-vocalization and, at the same time, aesthetic feelings, interest, curiosity, pleasure and self-reflection (Jacobs, 2011). In four studies, Miall and Kuiken (1994) collected segment by segment reading times and ratings from readers of three different short stories which contained a variety of foregrounded features (i.e., phonological, grammatical and semantic). They showed that the degree to which foregrounding is present in the segments of a story is a predictor of both reading times and readers' judgments of 'strikingness' and 'affect' and, most interestingly, that the effective foregrounded elements were primarily phonological and semantic.

In the context of the present work, these promising findings on the role of foregrounded elements in shaping the aesthetic and affective impact of a text can be applied to the phonological level of analysis in order to develop methods and tools for measuring the sublexical affective tone of texts. Thus, while developing such methods, a central question that I will address is *how to define a phonological foregrounded unit in a text*. Furthermore, the contribution of each extracted phonological unit to the sublexical affective tone needs to be addressed. This will be defined as a function of i) the affective potential of the phonological unit itself, and ii) the extent to which the unit is foregrounded.

3.5.3 Panksepp-Jakobson-Hypothesis

Another unique characteristic of the NCPM is its connection to the core affect system described by Panksepp (1998) on the one hand, and the poetic function of Jakobson's model (1960) on the other hand. This forms a novel hypothesis, termed the *Pankespp-Jakobson-Hypothesis* (Jacobs, 2011), which attempts to bridge the gap between neurobiological theories of emotion and complex linguistic models. The hypothesis states that since there was no time for the human brain to develop proper affective systems specifically for literary and other art reception in our evolution, emotional and aesthetic experiences evoked by this type of stimuli are processed in the same ancient affect circuits that we share with other mammals (Jacobs, 2011, 2015a).

So far, the hypothesis has received considerable empirical support that has shown the activation of core affect systems in response to short stories with negative emotional content (caudate body, left amygdala, see Altmann, Bohrn, Lubrich, Menninghaus, & Jacobs, 2012b) or to disgusting words (anterior insula, see Ponz et al., 2014; Ziegler et al., 2018). An additional statement that the hypothesis makes—related to the dual route of reading—is that background reading tends to be driven by the core affect system of fear, care, and anger, whereas the model's lower route is more associated with lust, play, and seek.

In regard to the present dissertation, the Panksepp-Jakobson-Hypothesis can be used to make predictions about the neural processes underlying the affective sound of words: The notion of affective iconicity proposes that the sound of words can possess iconic characteristics based on their phonological and acoustic representation. Since words are considered rather new achievements in the course of the evolution, the Panksepp-Jakobson-Hypothesis predicts that the affective potential of the sound of words will be processed in the same (subcortical) brain structures that we share with other mammals to evaluate other types of affective sounds, such as nonverbal emotional vocalization (Figure 3.3).

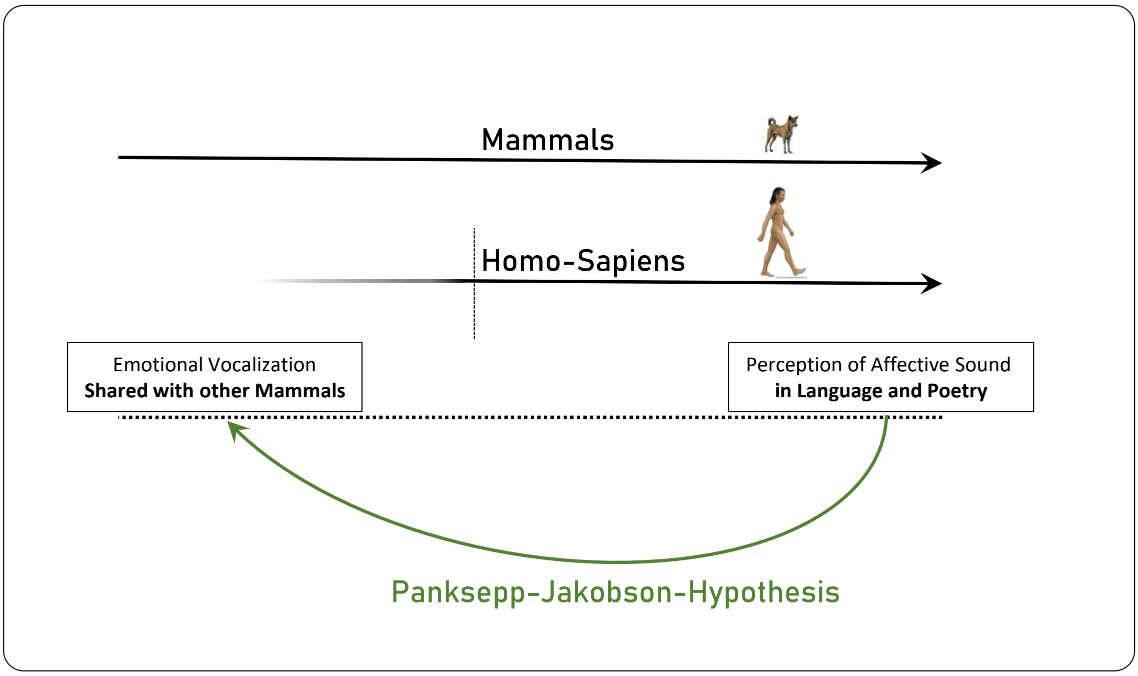


Figure 3.3. Prediction of the neural correlates underlying the affective potential of the sound of words as made by Panksepp-Jakobson-Hypothesis.

3.6 Literary Text as a 4x4 Matrix

In addition to the hypotheses formulated by the NCPM, the model provides descriptive tools that combine different text levels (metric, phonological, morpho-syntactical, semantic) with different groups of features (sublexical, lexical, interlexical, supralexical), forming a 4 × 4 matrix (Jacobs, 2015a) as displayed in Figure 3.4.

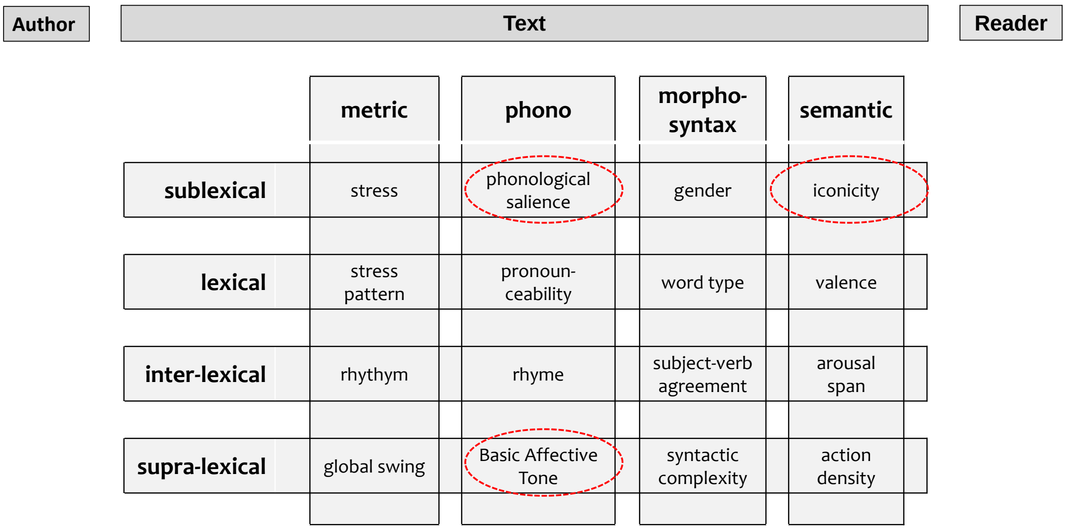


Figure 3.4. Illustration of the 4 × 4 matrix regarding the four text features (rows) combined with four different text levels (columns), with one example for each feature. The main focus of the present work is on iconicity, phonological salience, and basic affective tone (red dashed line).

Inspired by this method of text analysis, a number of studies have focused on different levels of analysis and different cells of the matrix. For instance, information on different lexical features (e.g., arousal rating) provided in databases such as BAWL (Võ et al., 2009; Võ, Jacobs, & Conrad, 2006), DENN-BAWL (Briesemeister, Kuchinke, & Jacobs, 2011a), ANGST (Schmidtke, Schröder, Jacobs, & Conrad, 2014), and their corresponding interlexical measures (e.g., arousal-span) have been used to quantify the emotion potential of texts (e.g., Jacobs, 2015b; Jacobs & Lüdtke, 2017; Lehne et al., 2015). Other studies have taken a multi-level approach and examined the role of features from different levels (Hsu, Jacobs, Citron, & Conrad, 2015a; Jacobs, Lüdtke, Aryani, Meyer-Sickendieck, & Conrad, 2016; Ullrich, Aryani, Kraxenberger, Jacobs, & Conrad, 2017). In contrast to the lexical and interlexical level, quantifying relevant variables at the supralexical level still presents big challenges because of the lack of similar databases or lists that could provide the relevant information. Here, in order to operationalize narrative structure and complexity, the use of appropriate tools from quantitative narrative analysis (QNA) or advanced qualitative-quantitative narrative analysis (Q2NA, see Jacobs, 2018b) has been suggested. The combination of these tools with machine learning algorithms to extract important features relevant for the effect in question presents a promising method for future research in this field. This has been demonstrated by recent studies (e.g., Jacobs & Kinder, 2018; Jacobs, 2017; Jacobs & Kinder, 2017).

With regard to the present work, the use of the 4 × 4 matrix clarifies which feature at which level needs to be addressed. The relation between sublexical features and the semantic level of text in affective domain, i.e., affective iconicity, is the primary focus of this work. Furthermore, in order to explore the effect of phonology at the supralexical level (i.e., basic affective tone) the foregrounded units at the phonological level (i.e, phonological salience) need to be detected and extracted from the text.

But how can sublexical structures influence emotional reactions to texts? Can sounds have intrinsic, autonomous meaning, particularly in literary and poetic language?

3.7 The Role of Sound in Literary Reception

The German linguist and philologist Eduar Sievers (1850-1932) was one of the first scholars in the 19th century who attempted to develop methodological tools to account for the acoustic, rhythmical and prosodic properties of verse in poems. He defined poetry as an auditory and acoustic phenomenon with sound as its organizing principle (Sievers, 1912). In his famous model of *Ohrenphilologie* (auditory philology), Sievers developed a systematic study of the psychophysiological conditions determining the production of human speech and the sound layer of language. His method of sound analysis (*Schallanlyse*) inspired other scholars, such as Gustav Becking (1894-1945), who combined the rhythmic patterns in music with physical movements (*Beckingkurven*), and Jakobson, who converged Sievers' and Becking's approaches with his own structuralist conception for the analysis of language and poetry (Jakobson, 1970; see Flack, 2016, for a review).

But the first and most influential schools that examined the role of sound in literature and poetry in a systematic and an empirical fashion were Russian Formalists and Prague Structuralists (Jakobson, 1960; Shklovsky, 1990; Trotsky, 1957; Mukarovsky, 1976). They took the position that the phonological structure of poetry has a function beyond the decorative, and should be an object of study in its own right. In their examination of sound in poetry, the Formalists went a long way towards answering the questions of how sound patterns are organized in verse and what onomatopoetic types of sound patterns may be identified. Their approach provided a procedural basis for subsequent investigations of sound in poetry using objective criteria and a linguistic focus (see Mandelker, 1983 for a review).

Despite the brilliant heritage of the study of sound in poetry from the Russian Formalists and Prague Structuralists, this field of research has not provoked much interest or many empirical investigations into language and literature in the past few decades. Besides a meagre selection of empirical works that I will review in the next chapter, research on this topic has often been of low repute due to methodological and theoretical shortcomings. With this dissertation, I aim at building upon the findings and insights of previous work in an attempt to address the limitations and issues evident in this line of research.

3.8 Conclusion

By addressing the potential neurocognitive and aesthetic-affective mechanisms underlying the processing of literary reading, the NCPM goes one step further than Jakobson's language model and provides a framework for empirical investigation in this field of research (see Table 3.1). Building upon Jakobson's model of language, the NCPM not only covers the necessary topics and features relevant to this work (as discussed in Chapter 2), but also makes testable predictions for investigating the neuronal and cognitive-affective bases of literary reading, which are used for the present dissertation.

Model	Accounting for			
	Iconicity	Emotion	Literature	Empiricism
Saussure	–	–	–	–
Pierce	✓	–	–	–
Bühler	✓	✓	–	–
Jakobson	✓	✓	✓	–
NCPM	✓	✓	✓	✓

Table 3.1. The NCPM compared to other models and theories (see Table 2.2) also accounts for empirical investigations of literary reading.

In order to reformulate my research questions and hypotheses in the most precise way and consider the limitations of previous research, in the next chapter I will review the empirical findings on affective iconicity at different levels of language for different semantic domains, and then proceed to present the corresponding empirical studies that were conducted over the course of the dissertation project.

Chapter 4

Empirical Findings on Iconicity

Iconicity is the existence of a similarity between some aspects of a word's form and some aspects of its meaning. A large number of empirical works in this field have used the term "sound symbolism" to refer to the same phenomenon. However, the use of the term "sound" in these studies does not necessarily refer to the acoustic representation of the word. Rather, it most frequently refers to the implicit sound of words emerging from the process of phonological decoding (Braun et al., 2009; Breen, 2014; Ziegler & Jacobs, 1995, see Chapter 3 for a detailed discussion). In the present work, as I will discuss in the next section, I use the term iconicity, originating from Peirce's theory of semiotics (see Chapter 2), to refer to any perceptual similarities between the sound (implicit or explicit) and meaning of linguistic signs.

4.1 Different Types of Sound-Meaning Mapping

Due to the existence of different types of sound-meaning mappings, many scholars have attempted to distinguish between the different motivations which cause such sound-meaning mappings in language. Wundt (1900) differentiated sound-imitation (*Lautnachahmung*) from sound-gesture (*Lautgebärden*) with the former representing the direct imitation of the sound through speech sounds (e.g., Tick-Tack) and the latter representing the sound accompanying humans' activities (e.g., mama). For the present work, I rely on current proposals (e.g., Dingemanse et al., 2015) that categorize the motivations for such sound-meaning mappings into two categories: iconicity and systematicity.

4.1.1 Iconicity

Iconicity is based on perceptual similarities between form and meaning. Onomatopoetic words (e.g., ‘cuckoo’, ‘splash’) constitute the most instructive examples for this category. In addition to onomatopoeia, a word can be considered iconic if its phonetic features or acoustic properties evoke sensory (including visual and tactile), motor, or affective experiences by systematically relating properties of such experiences. This is evident in ideophones (e.g., ‘twinkle’) or in mimetic words—as I discuss later—, or in affective responses associated with the phonology of swear words (Bowers & Pleydell-Pearce, 2011). The psychological aspect of the process of language comprehension in the iconic case can be “conceived of as the vicarious experiencing of events in the real world” (Segal, 1995; Zwaan et al., 2001; Zwaan & Yaxley 2003), or based on common neural coding across distinct sensory modalities (Dingemanse et al., 2015; Revill, Namy, DeFife, & Nygaard, 2014; Spence, 2011).

4.1.2 Systematicity

Systematicity is another type of sound-meaning relationship in language which is based on statistical regularities in language that link specific patterns of sound to specific semantic or grammatical concepts (Christiansen & Monaghan, 2016; Farmer et al., 2006). Calling up the Peircean trichotomy of different types of signs, systematicity is what is called an *index* in this model. That is, in systematicity there is no similarity or inherent relationship between sound and meaning but rather the over-proportional appearance of a specific sound and a specific category of meaning provides a hint for such an inference for other unknown words (just as smoke being an *index* of fire). Therefore, in this case the relation is not between a single word and simple referential meaning but rather between a large number of words and a limited number of abstract categories (Dingemanse et al., 2015). The relationship between word length (short vs. long) and the concreteness vs. abstractness of the word’s meaning (Reilly, Westbury, Kean, & Peelle, 2012) is one example of such non-arbitrary links.

Crucially, in this dissertation, my focus is on the phenomenon of iconicity, and an investigation of the role of systematicity in language is not within the scope of this work.

4.2 Iconicity in Different Languages

Some languages possess a large iconic lexicon that goes beyond onomatopoeia. Berlin, Hinton, Nichols, and Ohala (1994) provided an overview of a variety of languages with iconic elements. In sub-Saharan African languages, iconic words are referred to as *ideophones* (Childs, 1994) whereas in South-East Asian languages they are termed *expressives* (Diffloth, 1972; Enfield, 2005). However, the language with one of the most well-established and researched iconic systems is Japanese. Iconic words in Japanese have a grammatical class of their own, referred to as mimetics, which can describe a number of concepts such as vision, touch, motion, and emotion. This word class is relatively large with almost 4500 mimetics listed in the dictionary of mimetic words (Hamano, 1998).

Indo-European languages are assumed to have a small number of iconic words (Perniss, Thompson, & Vigliocco, 2010). Besides ononomatopoetic words, phonaesthemes are probably the most well-known example of iconic mappings in these languages. However, a debate on the actual type of the relationship between sound and meaning (i.e., iconicity vs. systematicity) for this category of words is still open. Phonaesthemes are sounds, sound clusters or sound types that are directly associated with a lexical category or meaning. The initial cluster /gl/ is often cited as an example of an English phonestheme. It occurs in many words used for 'shiny things': glisten, gleam, glint, glare, glam, glimmer, glaze, glass, glitz, gloss, glory, glow, and glitter (Wallis, 1699; Magnus, 2001; Bergen, 2004). Likewise, in German, nouns starting with /kno/ and /knö/ are mostly small and round: *Knoblauch* 'garlic', *Knöchel* 'ankle', *Knödel* 'dumpling', *Knolle* 'bulb', *Knopf* 'button', *Knorren* 'knot', *Knospe* 'bud', *Knoten* 'knot' (Magnus, 2001). By conducting a lexical decision task using phonaesthemes, Bergen (2004) showed that words were recognized more quickly when the prime and the target shared a phonaestheme, compared to when the prime and the target were semantically or phonologically related. These results suggest that words containing phonaesthemes appear to have a representation of their own in the mental lexicon.

Recent studies have sought to find universal patterns across different languages in the world. The most comprehensive study in this line to date is a study conducted by Blasi et al. (2016), who focused on 100 basic vocabulary from 6895 different languages and found similar associations between semantic domains and specific kinds of human speech sounds across these languages. For instance, the vowel /i/ is widely associated with small size, and /r/ with roundness, /n/ appears more in words for nose, /l/ for tongue, and /m/ in words for mother or breasts.

4.3 Iconicity and Different Domains

Psychological research into iconicity first flourished in the late 1920s. In his seminal study, Sapir (1929) raised the issue of whether phonemes in isolation are symbolic of differing size. He used two nonsense words (i.e., pseudowords) MAL and MIL and asked subjects to ascribe these two pseudowords to a large and a small table. Subjects consistently judged MIL to refer to the small and MAL to the large table. Almost at the same time, one of the most prominent types of iconic mappings was introduced by the German psychologist and phenomenologist Wolfgang Köhler (1887-1967) in the context of Gestalt psychology. In a series of psychological experiments, Köhler (1929) presented participants with angular and rounded shapes and asked which shape was called “takete” and which was called “baluba” (“baluba” was changed to “maluma” in Köhler (1947) to make it less similar to the word balloon). Participants matched angular shapes to takete and rounded shapes to baluba. These results were confirmed using the words “bouba” and “kiki”. 95% of adults preferred to label the rounded object bouba and the angular object kiki (Ramachandran & Hubbard, 2001). This iconic mapping between specific phonological forms and specific shapes has been shown to be detectable in children as young as 2-years-old (Maurer, Pathman, & Mondloch, 2006).

Another domain investigated in iconic research is that of motion, with the majority of research in this domain focused on Japanese mimetic words. For instance, Imai et al. (2008) presented participants with novel words created on the basis of existing Japanese mimetics and videos of various manners of walking. English-speaking and Japanese-speaking adults rated the extent to which the novel words and the actions matched similarly, indicating an iconic mapping between the sound of mimetics and their meaning based on general motor experiences.

Also, in the affective domain of pain, a single study on iconicity has been conducted in a similar fashion by using Japanese mimetic words. Iwasaki, Vinson, & Vigliocco, (2007a) presented English-speaking and Japanese-speaking participants with Japanese mimetic words for pain and asked them to rate the words on various semantic dimensions (aching, bothering, continuous, affecting wide areas). Again, the pattern of rating was similar between Japanese and English speakers, providing further evidence for the potentially universal effect of the phenomenon.

Inspired by the Sapir's study, the domain of size has attracted a number of researchers who attempted to find a direct mapping between some phonological and acoustic features (mostly for vowels) and the size of objects. For instance, Newman (1933) extended Sapir's categories of /i/ and /a/ to articulation point in the vocal tract for both consonants and vowels. He concluded that more frontal phonemes relate to smallness and vice versa. Later, Taylor and Taylor (1965) revealed statistically reliable relations within Newman's data of smallness with more frontal sounds (e.g., consonants /n/,/t/; vowels /e/,/i/) as well as largeness with more posterior sounds (e.g., /g/,/k/; /o/,/u/). However, the underlying mechanisms for this mapping remained a matter of debate. Recently, Thompson & Estes (2011) demonstrated that this effect follows a graded function in adults, and concluded that an acoustic mechanism based on the correlation between general physical size and vocal tract size (as proposed by Ohala's (1994) frequency code) can better account for the results than statistical learning. This idea is supported by the study of Peña, Mehler, & Nespor, (2011) who showed that even 4-month-old infants associate high frontal vowels with small object size and low posterior vowels with large object size.

This finding illustrates that an association based on prior knowledge, as postulated by statistical learning, is less likely than the inborn acoustic mechanism proposed by other theories such as the frequency code (Ohala, 1997; Ohala, 1994). I will return to this theory shortly.

Taste is another sensory domain in which iconic mappings have been found. Simner, Cuskley, & Kirby, (2010) presented participants with sweet, sour, bitter and salty flavored solutions in low, medium and high potencies alongside four different sound continua of F1, F2, voice discontinuity and spectral balance. Participants consistently matched lower F1 and F2 frequencies (approximating higher, more back vowels) with sweet flavors and higher F1 and F2 frequencies (approximating lower, more front vowels) to sour flavors.

Reviewing studies that focus on iconic mapping in different domains, a question remains as to whether iconicity is domain specific, or whether these domains are just examples of a more abstract type of iconicity (cf. Kantartzis, Kita, & Imai, 2011). For instance, size iconicity may be an example of a more basic semantic category, such as emotion. The "frequency code" theory by John Ohala (1994; 1997) provides a good example of such a possible link between intonational communication of affect and in iconic vocabulary: whereas high pitch sounds signify smallness, a non-threatening attitude and the desire for goodwill of the receiver, and a generally positive emotional valence, low pitch sounds convey largeness, threat, self-sufficiency, and a generally negative emotional valence. Regarding this categorization, vowels with a lower pitch (e.g., /u/, /o/, /a/) might be associated with largeness and negative emotion while vowels with a higher pitch (e.g., /i/, /e/) may connote smallness and positive emotion.

4.4 Facilitative Effect of Iconicity on Language Processing and Vocabulary Learning

Accumulating evidence from recent studies has shown iconicity to structure vocabulary in a supplementary way (Perniss & Vigliocco, 2014; Vigliocco & Kita, 2006) and to play an important role for both phylogenetic language evolution (Imai &

Kita, 2014; Perniss & Vigliocco, 2014; Roberts et al., 2015) and ontogenetic language development (Imai & Kita, 2014; Monaghan et al., 2014) by fulfilling the need to map linguistic form to human (sensory, motor and affective) experience (Perniss et al., 2010, 2014).

In addition to such direct acoustic mappings, such as in the case of onomatopoeia, iconic words can also evoke other sensory (including visual and tactile), motor, or affective experiences by systematically relating properties of such experiences to phonetic features or acoustic properties of spoken words. Iconic words may therefore be capable of directly activating the semantic domain that they refer to by bridging the gap between linguistic form and human (sensory, motor and affective) experience. From this perspective, iconicity may provide additional mechanisms for both vocabulary learning and language processing by means of direct sound-meaning mappings in neural systems devoted to perception, action and affective experience. This mechanism can potentially realize the embodiment of language (Vinson et al., 2015; Perniss & Vigliocco, 2014; Meteyard et al., 2015) and provide a new solution for the symbol grounding problem (Harnad, 1990, see also Barsalou, 1999; Glenberg, 1999).

Empirical evidence for both children and adults supports the idea that there is advantage of iconicity for learning the vocabulary of a language in which they have no experience. For instance, Japanese iconic verbs in the domain of motor actions have been shown to be learned better by English-speaking children, and their meaning could be more simply generalized than non-iconic words (Kantartzis et al., 2011; Imai et al., 2008). Also, adult speakers seem to be sensitive to consistency in form-meaning mappings, and can match the meaning of unknown iconic words from a foreign language more quickly and more accurately over learning blocks than for arbitrary words (Nygaard, Cook, & Namy, 2009). These results are in line with the analyses of longitudinal diary data which suggest that over the course of language development iconic words are in general acquired earlier, and potentially employed by infants as a bootstrapping mechanism on both lexical and phonological levels (Laing, 2014;

Monaghan, Shillcock, Christiansen, & Kirby, 2014; Thompson, Vinson, Woll, & Vigliocco, 2012).

As in the case of vocabulary learning, iconicity can facilitate language processing through alternative (and additional) links between linguistic form (visual or acoustic) and sensory, motor, or affective experiences (Meteyard, Stoppard, Snudden, Cappa, & Vigliocco, 2015; Perniss & Vigliocco, 2014; Vinson, Thompson, Skinner, & Vigliocco, 2015). Empirical support for this idea comes, for instance, from a study based on Köhler's kiki-bouba-effect. In a lexical decision study, Westbury (2005) used an implicit interference design with words and nonwords matching Köhler's stimuli's consonant characteristics. Stimuli were presented simultaneously to either congruent or incongruent round or angular shapes. Results showed an interaction between shape forms and nonword phonology, suggesting sound-shape mappings influence online language processing. Further support on the role of iconicity in lexical processing comes from research on sign languages, in which iconic relationships between form and meaning are far more prevalent than in spoken languages (Elliott & Jacobs, 2013; Taub, 2001). Vinson et al. (2015), for instance, could show that iconicity in British Sign Language facilitated picture–sign matching, phonological decision, and picture naming, suggesting that during lexical processing iconic words benefit from an additional path between form and meaning by activating conceptual features related to perception and action (see also Thompson et al. 2012).

4.5 Iconicity, Affect, and Poetry

As I outlined in Chapter 3, the role of emotion and affective experiences has been neglected to a surprising extent in research on iconicity. At the lexical level, there have been a few studies focused on iconic mappings in the affective domain with the study of Zajonc, Murphy, & Inglehart (1989) representing one of the first empirical works in this line of research. Taking an embodied approach, they examined the relationship between some vowels and the specific emotion related to the facial muscles needed for the articulation of those vowels. By contrasting the umlaut /ü/ with other vowels, Zajonc et al. hypothesized that facial muscle feedback from the corrugator muscle associated with its production would cause rather negative affective states. Pleasantness and mood ratings of American and German subjects became indeed more negative after the utterance or after reading stories with higher occurrence of this specific vowel (Zajonc et al., 1989). In another study, Wiseman and van Peer (2003) focused on the articulatory aspects of consonant, and revealed that when German and Brazilian participants were asked to produce fantasy words corresponding either to the emotions experienced at a wedding or at a funeral, they tended to use similar consonants for respective emotional states, and this was independent of their native language. While nasal sounds (/m/, /n/) were more frequently used for the expression of sadness (funeral), plosive sounds (/p/, /b/, /d/, and /t/) were better suited to the expression of happy feelings (wedding) (Wiseman & van Peer, 2003).

At the text level, one of the pioneering empirical works on the iconicity in the affective domain is the study conducted by Fónagy (1961) in which he compared the distribution of phonemes in poems of the Hungarian poet Petöfi. The poems were categorized into two groups of 'aggressive' and 'tender' based on normative ratings. He found that 'aggressive' poems contained a greater proportion of voiceless plosives such as /t/ and /k/ whereas 'tender' poems predominantly included more sonorants such as /m/, /n/, and /l/. The fact that one of the earliest empirical works on a sound-emotion relation was devoted to poetry should not be surprising. Indeed, the literary genre of poetry has been of particular interest for the investigation of iconicity much more often than the lexical level (e.g., Albers, 2008; Auracher et al., 2010; Jakobson & Waugh, 1979/2002; Fónagy, 1961; Schrott & Jacobs, 2011; Tsur, 1992). Poetry is deeply rooted at the aesthetic and perceptual level in the domains of speech and sound (Jacobs et al., 2015; Jacobs & Kinder, 2015; Schrott & Jacobs, 2011; Wolf, 2005). Emphasis on sublexical units such as phonemes or syllables through stylistic devices such as onomatopoeia, parallelisms, or alliterations may provide good examples of the importance of sound aspect in poetry. At the same time, the expression and elicitation of emotions lie in the basis of poetry (Lüdtke, Meyer-Sickendieck, & Jacobs, 2014; Meyer-Sickendiek, 2011; Winko, 2003). The interaction of sound and affect in lyrics and poems has been shown, for instance, in the effect of meter and rhyme on aesthetic appreciation, intensity of processing and emotional perception (Menninghaus et al., 2014; Obermeier et al., 2013; Bohrn et al., 2012b, 2013).

While a more comprehensive review of previous empirical works in this field can be found in the theoretical sections of the studies in the empirical part, I will review a few more influential studies on the topic. In an attempt to find some universal patterns, Albers (2008) conducted a comparative analysis of Old Egyptian hymns and lamentations together with hymns and ballads by Johann Wolfgang von Goethe. Plosive sounds were found to occur significantly more frequently in hymns of both sources, whereas nasals were more frequent in lamentations and ballads (Albers, 2008). Similarly, a cross-linguistic study conducted by Auracher et al. (2010) reported a higher frequency of the plosives in poems rated as happy versus a higher frequency of nasals in poems perceived as sad; with consistent results for German, Chinese, Russian and Ukrainian participants and poems (Auracher et al., 2010). Some other studies exclusively focused on the works of a single author. For instance, Miall (2001) compared passages from Milton's “Paradise Lost” that either dealt with depictions of Hell or Eden. Passages about Hell were found to contain significantly more front vowels and hard consonants than passages about Eden which contained more medium back vowels.

Last but not least, Whissell's pioneering analysis of *phonoemotionality* (e.g., Whissell, 1989, 1999, 2000, 2011) is presumably one of the most comprehensive works in this field. In several analyses of English poetry and lyrics, Whissell proposed that most of the basic sounds of English have emotional connotations attached to them (see, for instance, Whissell, 2000). She created the '*Dictionary of Affect in Language*' (Whissell, 1989) in order to validate these connotations. Consequently, she found that, for example, the /l/ sound has positive and gentle connotations while /r/ and /g/ sounds are harsh and unpleasing. By further analysis of these findings, Whissell (1999, 2000) attributed phonemes' emotional quality to both place and manner of articulation as being variably related to different positions in the affective space (e.g., pleasantness, sadness, passivity, etc.). In addition, she analyzed the phonological material of the poetic works of Edgar Allan Poe (Whissell, 2011) reporting that Poe used “pleasant, sad, and soft sounds” more frequently than sounds that were categorized as “active”.

4.6 Iconicity and Brain Imaging

At the level of neural substrates, only a limited number of neuroimaging studies have investigated the underlying mechanisms associated with the processing of iconic words. These have focused primarily on onomatopoeia and Japanese mimetic words. Results suggest that onomatopoetic words activate the relevant cortical areas that they are associated with. For instance, Japanese mimetic words expressing laughter and pain have shown to activate the premotor brain areas associated with an actual laughter and pain, and in addition, striatal reward area and cingulate cortex, respectively (Osaka et al., 2003; Osaka, Osaka, Morishita, Kondo, & Fukuyama, 2004). However, since embodiment theories assume that arbitrary words also activate relevant domain-specific sensorimotor areas (Hauk, Johnsrude, & Pulvermüller, 2004; Vigliocco et al., 2009; Zwaan, 2004), the potential advantage in processing of iconic words remained unclear (cf. Lockwood & Dingemanse, 2015). Kanero et al. (2014) overcame this shortcoming by comparing onomatopoetic expressions that were related to motion and shape with arbitrary words from the same semantic domains. Greater general activation and a cluster of activation in the right posterior superior temporal sulcus (pSTS) was observed for onomatopoeic words compared to their arbitrary counterparts. Based on this finding, the authors suggested that iconic words, in addition to lexical processing, profit from a sublexical (i.e., sound) processing network; with the pSTS working as a hub for integration of multimodal (i.e., lexical and sublexical) information. This view is in line with previous findings showing a greater activation for bimodal information (e.g., onomatopoetic words imitating animal calls) in the left and right superior temporal sulcus (STS) than for unimodal information (e.g., either animal names, or animal calls; Hashimoto et al., 2006).

By extending the word material to a multi-language stimulus set, Revill et al. (2014) provided further support for the advantageous processing of iconic words, and for the potential role of areas engaged in multimodal sensory integration beyond those involved in semantic processing. They asked English speakers to choose the meaning of foreign words that were either iconic or non-iconic from one of four corresponding antonym pairs. Iconic words were matched with the consistent meaning at above the rate of chance, and were associated with an enhanced BOLD signal in the left superior parietal cortex known to be involved in multisensory integration.

In general, these data point to the possible additional processing network for iconic words which may consist of more direct links between information from semantics and from phonological forms with corresponding neural hubs as a convergent structure for integration of information. If so, iconic words would be expected to be more immune to neurological damage that affects language processing network, as in aphasic patients, for example. In fact, in a recent lesion study involving individuals with aphasia following left-hemisphere stroke (Meteyard et al. 2015) a consistent processing advantage was seen for onomatopoetic words in reading aloud and auditory lexical decision; two tasks that rely on mapping between semantics and phonology.

4.7 Conclusion

Taken together, the review of findings of previous work on iconicity provides valuable insights into the existence, the psychological reality, and the possible explanations of this phenomenon. On the other side, and of particular relevance to the affective domain, results of previous work suggest a large hidden potential and a number of unanswered questions that require exploration in this field. Besides the limitations of assessing the affective potential of sound in language, which I will discuss in the next chapter, a lack of empirical research in the experimental setting—with a systematic manipulation of independent variables—seems to be a common characteristic of previous work on affective iconicity. Also, the psychological and neural mechanisms underlying the effect of affective sound on affective meaning would appear to remain largely unknown. All this has contributed to the motivation of my research questions, which I will discuss in the next chapter.

Chapter 5

Research Objectives

In this chapter, I will summarize the limitations and shortcomings of the previous studies reviewed in Chapter 4. Based on this, I shall then formulate the precise research questions that I aim to answer in the context of this dissertation by means of empirical investigations. I will introduce these empirical investigations at the end of this chapter.

5.1 Limitations of Previous Research

The majority of previous work in this field of research has focused on a limited number of semantic domains usually related to the sensorimotor information (e.g., size, or motion). As I will argue, this restricted approach to the operationalization of meaning is presumably the most important issue of the previous work on iconicity. Similar to the earlier model of language processing that focused on cognitive processes alone (see Chapter 2 and 3), only a few studies have focused on the contribution of emotion and affective processes to sound-meaning mappings. This neglect of research on the role of emotion in iconicity is surprising, given that the expression and perception of affective states are fundamental aspects of human communication (Darwin, 1871; Wundt, 1908) that have been proposed as the original impetus for language evolution (Darwin, 1871; Panksepp, 2010). As stated in the following by Darwin, and echoed by other scholars (cf. Christiansen & Kirby, 2003), the close relationship between emotion and sound probably lays the groundwork for evolution of language:

"I cannot doubt that language owes its origin to the imitation and modification, aided by signs and gestures, of various natural sounds, the voices of other animals, and man's own distinctive cries [. . .] serving to express various emotions, as love, jealousy, triumph, and serving as a challenge to their rivals. The imitation by articulate sounds of musical cries might have given rise to words expressive of various complex emotions." (Darwin, 1871, p. 56)

Since iconicity has been considered a factor underlying language evolution (Carolis, Marsico, & Coupé, 2017; Perniss & Vigliocco, 2014; Ramachandran & Hubbard, 2001; Reilly et al., 2012; Schrott & Jacobs, 2011; Thorndike, 1943), the effect of iconicity in today's language must be most evident in the communication of affect and in the relationship between the potential affectivity in the sound of words and their affective meaning. Hereby, iconicity can serve as an interface for accomplishing the need to map linguistic form to human affective experience as a vital part of meaning making.

Connected to the previous issue, there has been a lack of proper operationalization of meaning in the previous work. By focusing on the affective domain, in this work, affective meaning of words was measured according to the dimensional view stemming from the tradition of Wundt's three factor theory (Wundt, 1922), i.e., valence, activation/arousal, and potency. These factors were also identified as three underlying core dimensions accounting for the majority of variance in the seminal factor analysis conducted by Osgood (Osgood, 1952; Osgood et al., 1957; see also Russel and Mehrabian, 1977). For the rating of words' affective meaning, I relied on empirically approved measures of valence and arousal, which represent prototypical emotional episodes in dimensional theories of emotion, called core affects (e.g., Russel, 2003; Barrett, 2006). These allow for a basic and, potentially, the most relevant distinction between different concepts. For this, I used the Berlin Affective Word List (Võ et al., 2009; BAWL) as a normative database containing a representative mass of 2694 German words that have been cross-validated in (to date) over 50 empirical studies regarding experiential, behavioral, and neurobiological levels of analysis (Jacobs et al., 2015).

The next issue that prevented previous work from identifying the degree of iconicity in a word—in order to use it in carefully controlled experimental paradigms—has been the lack of a quantitative measure for assessing the sound of words. This limitation makes the identification of a similarity between sound and meaning a difficult, if not impossible, task. This is a general issue in the study of sound-meaning correspondences and has been mentioned in the work of many scholars, such as Fónagy (2001):

"The sound spectra of thunder could perhaps be compared with the acoustic pattern of words denoting 'thunder' in different languages. But how can one measure the resemblance between the sound shape of French glisser, English (to) glide, and the meaning of these words?"

That is, in order to measure the degree of iconicity of words and to quantitatively distinguish iconic words from non-iconic ones, an objective measure of the sound needs to be developed. This measure will then be used as an experimental factor to design appropriate experiments and to further investigate the effect of iconicity in language processing. In this work, similar to the measurement of the meaning, the perspective variables for the affective potential of the sound of words will rely on the same affective dimensions of valence and arousal.

As the review of the previous work clearly shows, the majority investigate the effect of iconicity for nonwords rather than real words. This has been motivated by the fact that natural words in a language are linked to predetermined semantic concepts that are automatically activated during word recognition (see Chapter 3). Therefore, in order to disentangle phonological from semantic effects, these studies used nonword stimuli (e.g., Parise & Pavani, 2011; Westbury, 2005). In other cases, when using real words, they focused either on onomatopoeia, and ideophones, including Japanese mimetic words (Dingemanse, Schuerman, Reinisch, Tufvesson, & Mitterer, 2016; Iwasaki, Vinson, & Vigliocco, 2007a; Kwon & Round, 2014;

Lockwood, Hagoort, & Dingemanse, 2016), or on cases typically considered as systemticity (Bergen, 2004; Farmer et al., 2006; Reilly et al., 2012). This restricted focus on nonwords and marginal cases in language practice (i.e., ideophones) surely represents a limitation for generalizing results to real words. In this work, I will focus on a large number of 'ordinary' words to investigate the effect of iconicity as a general phenomenon in language. By providing two quantitative measures for aspects of meaning and the sound of words (the last two issues), I will attempt to provide a method for measuring the degree of iconicity in any given word in lexicon. This will fulfill the necessary requirements that any word needs to take part in further empirical investigations.

The next issue pertains to the largely unknown role of the sound aspect in meaning making. That is, results of previous studies do not draw a conclusion whether the sound of a word has a standalone role in meaning making, and whether it influences other cognitive processes responsible for constructing the meaning of a word. Such an inconclusiveness of previous work about this potential effect lies, in great part, in other aforementioned issues, i.e., using of nonwords rather than real words, and a lack of quantitative measurements for both sound and meaning hindering a carefully controlled experimental design for such an investigation. Therefore, it is of utmost interest for this work to examine Saussure's model of language (Chapter 2) and to address the following open question: Is assigning meaning to words solely determined by arbitrary links between the signifier (sound image) and the signified (concept), *OR*, is it co-determined by inherent qualities of the signifier and by the percept derived from words' acoustic-phonetic features? This question can be divided into two questions. The first question concerns the effect of sound on the final evaluation of the words' affective meaning which will be examined in a rating study. If the sound of words has any contribution to the affective evaluation of words' meaning, a significant amount of variance of ratings will be accounted for by acoustic-phonetic features of the words. The second question will focus on the role of sound in online lexico-semantic processes. Here, I will investigate whether affective potential in the sound of a word can influence semantic decisions about the meaning of that word, and whether language users make use of such information in the sound to more readily access the meaning of words.

Although some initial studies investigated the neural mechanism potentially underlying the effect of iconicity in language processing, these works, in general, have encountered the same limitations as described above, i.e., the use of nonwords, or onomatopoeia as stimulus material. Exploring the neural correlates of the affective potential of the words' sound, on the one hand, and its interaction with the affective meaning of words, on the other, is, therefore, one of the main focuses of the present dissertation.

Lastly, investigations of iconicity at the level of whole text face the general problem of the lack of a statistical measure of the sound effect that can be experimentally validated and independently applied to any given text. A general problem of the studies in this field has been a focus on the mere frequency of occurrence of the phonemes of interest, which, as suggested by the foregrounding theory, could be misleading due to specifics of phoneme distributions in the poetic language mode.

5.2 Research Questions and Hypotheses

My research into the different processing layers and textual level of affective iconicity can be summarized in six precise research questions (in short RQ). In this dissertation, my colleagues and I conducted six different studies addressing each of the following:

<u>RQ 1</u>. Do speech sounds (i.e., phonemes) in words possess affective characteristics? Can they evoke affective responses at the perceptual level? And if so, what are the acoustic-phonetic features underlying such affective potential of the sound of words? (Study 1).

<u>RQ 2</u>. Can speech sounds evoke affective responses at the neural level? How is the affective potential of the sound of words represented in the neural network when language users listen to that word? Is the affective brain response to the sound of

words similar to other types of affective sounds (e.g., nonverbal emotional vocalization)? (Study 2).

RQ 3. Can the affective sound of a word implicitly influence language users when giving their affective judgments on the meaning of that word? (Study 1). Can the affective sound of a word, and in particular its congruence with the affective meaning (i.e., iconicity) influence semantic processing, and help language users to more readily access the meaning of that word (i.e., more quick and more accurate semantic decisions)? (Study 3)

RQ4. If affective iconic words are processed more quickly and more accurately than non-iconic words (RQ 3), what is the neural mechanism underlying this beneficial processing? Do iconic words profit from an additional processing network consisting of neural hubs that integrate affective information from different sources (i.e., sound and meaning)? (Study 4)

RQ5. How can salient, over-proportionally used phonemes, and phoneme clusters be extracted from a (literary) text, in order to build further methods and tools capable of measuring the affective tone of a text at the sublexical level—as suggested by the stylistic device of foregrounding, and the NCPM? (Study 5)

RQ6. How can the sublexical affective sound of a text (i.e., the *basic affective tone*) be measured by applying an adequate statistical operationalization? How can this measure be tested for a collection of (literary) texts in a theoretical model (i.e., NCPM) that consider not only the text, but also the role of readers and author? (Study 6)

Each of these questions is the subject of a research study that is presented in a separate chapter as described in the following, except for Chapter 6 (Study 1) which is intended to provide answers for RQ1 and a part of RQ3.

5.3 Conceptualization of the Empirical Part

Chapter 6 – (Study 1):

Why ‘piss’ is ruder than ‘pee’? The role of sound in affective meaning making

This chapter attempts to address the long-standing conjecture whether phonemes have any inherent semantic content or any contribution to words’ meaning beyond the conventional links in the language system. A series of statistical analyses and behavioral studies were conducted to provide an answer for this question. By focusing on the words in a large-scale normative database (i.e., BAWL-R; Võ et al., 2009), including ratings of affective meaning for both valence and arousal, we hypothesized that the phonological word forms can contribute to the ratings of words’ affective meaning, so that a statistically significant portion of the variance in ratings is associated with words’ phonology and can be explained by words’ acoustic-phonetic features. To test this hypothesis, we formulated a statistical model and attempted to extract and amplify the potential effect of phonemes on ratings (*Phonological Affective Potential*) through signal averaging. We synthesized the words in the database, and extracted a number of acoustic features from them, thereby examining whether the *Phonological Affective Potential* is linked to (and derived from) words’ acoustic features.

Further, we tested the association of the *Phonological Affective Potential* with words’ affective sound to ensure its affective nature. The affective sound of words was assessed through two different independent methods. Importantly, the psycho-acoustic model that we developed in this work provided us with a reliable statistical measure for the sound of words (cf. limitation of previous studies) and laid the groundwork for our next studies investigating the effect of sound on meaning.

Finally, in order to identify the perceptual cues underlying the effect of words’ sound on the evaluation of their affective meaning, we constructed acoustic profiles and separately tested the relationship between each of the affective acoustic features and the *Phonological Affective Potential* on one side, and the measures for affective sound on the other side. Revealing the perceptual acoustic cues which likely underlie the effect of sound on meaning, we performed further analyses to explore phonetic features (e.g., vowel length, voicing) potentially causing this effect.

Chapter 7 – (Study 2):

The Sound of Words Evokes Affective Brain Responses

This chapter pertains to the neural networks underlying the processing of the affective potential of the sound of words. Although the brain networks involved in emotion processing for both verbal and nonverbal stimuli have been well studied, the neural correlates of the affective potential of a word's sound is largely unknown. In this study, we examined the neuropsychological reality of sublexical sound effect, and attempted to identify its underlying brain network. To quantify the affectivity of the sound of words we used the psycho-acoustic model developed in Study 1. Since the model is based on similar acoustic features to those modulating emotional vocalization and affective prosody, we hypothesized that affectivity in the sound of a word would be processed in similar brain regions to those involved in processing other types of affective sounds. For this, we conducted an fMRI study involving a passive listening task, and presented participants with words varying in their *sublexical arousal* while controlling for *lexical arousal*.

Chapter 8 – (Study 3):

Affective Congruence between Sound and Meaning of Words Facilitates Semantic Decisions

This study intended to extend the scope of findings from the first study (Study 1). It asked whether a congruence between sound and meaning (iconicity) can facilitate evaluative decisions on words' affective content. The basic idea behind this question comes from the assumption that iconic words may be capable of directly evoking sensory, motor, or affective experiences by systematically relating properties of such experiences to phonetic features, or acoustic properties of words. Thus, iconic words may profit from additional processing networks that can facilitate lexical processing, and in particular semantic decision. To test this hypothesis, we focused on affective arousal, as the result of Study 1 emphasized its important role in iconic mappings. We decided on a 2x2 design with the sublexical affective sound (*sublexical arousal*) and the lexical affective meaning (*lexical arousal*) as two experimental factors. *Sublexical arousal* was measured by applying the psycho-acoustic model (developed in Study 1) to the acoustic features of the words in spoken form. For lexical arousal, we once again used the affective ratings from the BAWL-R. By organizing words in two groups of iconic and non-iconic, we asked participants (in a two-alternative forced choice task) to decide as quickly and accurately as possible whether the meaning of visually presented words was "exciting" or "calming". According to our alternative hypothesis, shorter latencies and higher accuracy were expected for iconic words compared to their non-iconic counterparts.

Chapter 9 – (Study 4):

Moving beyond Onomatopoeia: Affective Iconic Words Benefit from Sound-Meaning Integration in the Left Amygdala

In this chapter, I focused on the neural correlates of affective iconicity in language processing. This study investigates the underlying neural correlates and mechanisms likely responsible for the behavioral findings of Study 3. That is, we asked whether iconicity provides additional neural mechanisms for language processing by means of direct sound-meaning mappings in neural systems devoted to affective experience. In an fMRI experiment with a similar 2x2 design to Study 2, we compared the brain responses to iconic words with those to non-iconic words. We predicted a generally greater activation for iconic words, particularly in the brain areas associated with affective processing and in *convergence zones* that are responsible for supramodal representation of emotional information from different sources; i.e., phonological/acoustic information (related to *sublexical arousal*) and semantic information (related to *lexical arousal*). Suggested brain areas as candidates for supramodal integration of congruent emotional information are the (left) amygdala and the posterior cingulate cortex, in which we expected an enhanced BOLD signal for iconic words.

Furthermore, to investigate the interaction between the brain regions involved in iconic sound-meaning mappings, we performed a functional connectivity analysis using a "generalized psychophysiological interactions" approach (gPPI). Adopting two independent seed regions representative of the processing of sound (superior temporal gyrus) and meaning (inferior frontal gyrus), we hypothesized that iconicity significantly increases the coupling between these two seeds, on the one side, and the convergent zones integrating emotional information, on the other.

Chapter 10 – (Study 5):

Extracting salient sublexical units from written texts: 'Emophon', a corpus-based approach to phonological iconicity

The fifth study of this dissertation investigates the effect of iconicity at the text level. That is, it provides methods and tools for further investigations of the sound-meaning relationship at this supralexical level. A major shortcoming evident in previous approaches in this field has been focusing on the mere or rational frequencies of occurrence of single phonemes or classes of phonemes (see previous section). Inspired by the notion of foregrounding (see the NCPM, Chapter 4), we focused on deviant phonological units within a text, and attempted to deliver a measure of phonological salience based on the deviation of the observed frequency from the expected frequency of particular phonemes or phonemic clusters in a given text. To identify the salient units within a text, we developed a non-linear probabilistic model capable of predicting the level of salience for each phoneme or phoneme-cluster in the text. The basic idea behind the model is to weight the frequencies of occurrences of a sublexical unit in a text by comparison to a linguistic corpus (including 25 Million German words) serving as a reference. Implementing this model in a computational application, we provided a text analysis tool, called EMOPHON, which automatically delivers information about sublexical phonological salience. Next, we hypothesized that the numbers of salient phonological units—as provided by our tool—significantly differ between poetic and prosaic texts, and tested this hypothesis by comparing the results of analyses over 20 poems and 20 text passages from newspapers.

Chapter 11 – (Study 6):

Measuring the Basic Affective Tone of Poems via Phonological Saliency and Iconicity

Finally, the last study asks whether and to what extent the overall affective meaning of a text is (co-)determined by the use of specific phonological units. In this study, we examined the relation between the phonological inventory of the texts to both the author's affective categorization and readers' perception of the poems. For this, we developed a quantitative measure for assessing the affective sound of texts based on foregrounded phonological units and their iconic affective properties; termed *basic affective tone*. We combined our method for extracting the foreground elements from the text (Study 5) with the insights about the contribution of the affective quality of each single phoneme to the total rating value for each word (i.e., phonological affective value) obtained from Study 1. By doing so, we provided an independent statistical measure which can make predictions about the affective load of phonological structures in a single poem without the necessity of further comparisons of the results with other poems.

In order to test the method, we applied the novel measure to the volume of German poems "verteidigung der wölfe" (defense of the wolves) by Hans Magnus Enzensberger, who categorized these 57 poems as friendly, sad, or spiteful. Based on the NCPM, we focused on the three contributing factors of communication: that is, the author, the reader and the text. By choosing the poems of Enzensberger's "verteidigung der wölfe," which the author himself assigned to three affective categories (friendly, sad, and spiteful), we were able to incorporate considerations about the author as a factor for statistical analyses. On the recipient side, we conducted an extensive rating study to assess readers' judgments on three affective, author-based dimensions of friendliness, sadness, and spitefulness, together with the dimensions of valence and arousal. In a series of statistical analyses, we tested the relationships between our statistical measures of the basic affective tone, on the one hand, and both author-based emotion categories, and reader-based emotion ratings, on the other hand.

II

Empirical Part

Chapter 6

Why 'piss' is ruder than 'pee'?

The role of sound in affective meaning making[2]

6.1 Abstract

Most language users agree that some words sound harsh (e.g., grotesque) whereas others sound soft and pleasing (e.g., lagoon). While this prominent feature of human language has always been creatively deployed in art and poetry, it is still largely unknown whether the sound of a word in itself makes any contribution to the word's meaning as perceived and interpreted by the listener. In a large-scale lexicon analysis, we focused on the affective substrates of words' meaning (i.e., *affective meaning*) and words' sound (i.e., *affective sound*); both being measured on a two-dimensional space of valence (ranging from pleasant to unpleasant) and arousal (ranging from calm to excited). We tested the hypothesis that the sound of a word possesses affective iconic characteristics that can implicitly influence listeners when evaluating the affective meaning of that word. The results show that a significant portion of the variance in affective meaning ratings of printed words depends on a number of spectral and temporal acoustic features extracted from these words after converting them to their spoken form (study1). In order to test the affective nature of this effect, we independently assessed the affective sound of these words using two different methods: through direct rating (study2a), and through acoustic models that

[2] This chapter is published as: Aryani, A., Conrad, M., Schmidtke, D., & Jacobs, A. (2018). Why 'piss' is ruder than 'pee'? The role of sound in affective meaning making. *PloS one*, *13*(6), e0198430.

we implemented based on pseudoword materials (study2b). In line with our hypothesis, the estimated contribution of words' sound to ratings of words' affective meaning was indeed associated with the affective sound of these words; with a stronger effect for arousal than for valence. Further analyses revealed crucial phonetic features potentially causing the effect of sound on meaning: For instance, words with short vowels, voiceless consonants, and hissing sibilants (as in 'piss') feel more arousing and negative. Our findings suggest that the process of meaning making is not solely determined by arbitrary mappings between formal aspects of words and concepts they refer to. Rather, even in silent reading, words' acoustic profiles provide affective perceptual cues that language users may implicitly use to construct words' overall meaning.

6.2 Introduction

Human language has generally been considered to be entirely symbolic in that words convey meaning through conventional and arbitrary links to concepts they refer to (Saussure, 1916/2011). From this perspective, phonemes (i.e., the speech sounds that constitute words) have no inherent semantic content nor have they any stand-alone contribution to words' meaning. Nevertheless, even a naïve reader—without prior knowledge of such literary devices as cacophony or euphony—would experience how, for instance, in Poe's verse "*…Hear the loud alarum bells -- Brazen bells! -- What tale of terror, now, their turbulency tells!*" (Poe, 1881), the explosive consonant /t/ and other harsh and discordant sounds (e.g., hissing sibilants /s/ and /z/) evoke a feeling of "terror" provoked by "brazen" bells.

Within literary studies, many have noted that poetry achieves much of its affective aesthetic impact through sound manipulation, and that phonological structure has a semantic function beyond the decorative (Jakobson, 1960; Schrott & Jacobs, 2011; Tsur, 1992b). In a similar fashion, swear words usually possess specific phonological patterns that can potentially amplify the negative emotional response that they mean to evoke (Bowers & Pleydell-Pearce, 2011).

Looking at the famous seven words listed by American comedian George Carlin that “you can never say on television” (Carlin, 1990) reveals that all of these words contain voiceless stops (/t/ and /k/) or hissing sibilants (/s/ and /ʃ/), which are fortis consonants, articulated with greater oral pressure and relatively higher muscular force compared to their lenis counterparts.

However, despite the fact that influential linguists and experimental psychologists throughout the last century promoted the idea that the sound of a word may have a synchronic, productive effect on overall meaning construction (Jakobson, 1937; Kohler, 1947; Sapir, 1929), the notion of the arbitrariness of the linguistic sign (Saussure, 1916/2011) has generally dominated research on human language.

More recently, a growing body of research challenges the idea of absolute arbitrariness by providing evidence for non-arbitrary sound-to-meaning correspondences (see Dingemanse et al., 2015; Lockwood & Dingemanse, 2015; Schmidtke, Conrad, & Jacobs, 2014, for reviews) including some universal patterns across various languages of the world (Blasi et al., 2016). These results assign a supplementary function to sound-to-meaning correspondences that structure vocabulary (Perniss & Vigliocco, 2014; Vigliocco & Kita, 2006) and play an important role for both phylogenetic language evolution (Imai & Kita, 2014; Perniss & Vigliocco, 2014; Roberts et al., 2015) and ontogenetic language development (Imai & Kita, 2014; Monaghan et al., 2014). Nonetheless, despite the increasing number of studies examining sound-to-meaning associations, to the best of our knowledge, there has been no empirical study examining whether specific properties in the sound of a real word play a part in contributing to its overall meaning. With the present study, we aimed at addressing this research question. By focusing on the ‘affective meaning’ of words, and by providing reliable quantitative measures for ‘affective sound’ of words, we investigated how the sound of a word potentially contributes to its meaning as perceived and evaluated by the listener. A further goal of this study was to explore the affective acoustic cues and their underlying phonetic features that may implicitly influence language users when evaluating words' affective meaning.

6.2.1 Motivation for the present study

Our approach was motivated by a number of limitations evident in previous work. Experimental research based on behavioral data has hitherto merely investigated the links between some selective, rather isolated attributes of meaning (e.g., the physical size of the referent) and some aspects of sound (e.g., intrinsic pitch of vowels) mainly by using nonword stimuli (see Thompson & Estes, 2011) for supporting a graded relationship between sound and meaning, and (Ohala, 1996), for an evolutionary perspective on the phenomenon). Such approaches exhibit three major limitations that we aimed to address in the present study.

The first limitation relates to the focus on semantic effects of phonemes in nonwords instead of natural words. Such studies are motivated by the fact that natural words in a language are linked to predetermined semantic concepts that are automatically activated during word recognition. In order to disentangle the effect of phonology from that of semantics, the majority of previous studies therefore relied on nonword stimuli usually used in a forced-choice paradigm thus limiting the generalizability of the results to real words. For instance, the phonemes /ɑ/ and /ɪ/ when used in experimentally manipulated nonwords—as in "mal" and "mil" in the seminal study by Sapir (Sapir, 1929)—have repeatedly been suggested to denote big and small objects, respectively (Dingemanse et al., 2015; Schmidtke, Conrad, et al., 2014). However, in a natural language like English, they appear in the corresponding semantic concepts in the opposite way: /smɑl/ and /bɪg/. This begs the question to what extent the results of these studies can be linked to natural language processing and whether the assumed quality of phonemes has, if any, effects on the evaluation of meaning for real words.

A second issue relates to the problem of deciphering the likely cause(s) of sound-to-meaning correspondences. Proposals on non-arbitrariness of language distinguish between two types of motivations for such sound-meaning mappings (Dingemanse et al., 2015): Iconicity, which is based on perceptual similarities between sound and meaning (e.g., onomatopoeia), versus systematicity which is based on statistical regularities in language that link specific patterns of sound to specific semantic or grammatical concepts (Christiansen & Monaghan, 2016; Farmer et al., 2006). Besides some familiar and straightforward examples of iconicity—such as onomatopoetic words—research in this field still faces the question of whether existing findings on the relationship between sound and meaning are caused by specific distributions of phonemes in a language (i.e., systematicity), or by perceptual qualities that phonemes inherently convey (i.e., iconicity). The phonaestheme /sn-/ appearing as an initial sound cluster in many English words related to 'mouth' or 'nose' may serve to illustrate this dilemma (24). In this case, there has been no empirical support showing whether there is a specific (nasal) quality in the sound of /sn-/ that is linked with the concepts of 'mouth' or 'nose', or rather the organization of the vocabulary is designed in a way that this specific sound cluster over-proportionally appears in words that are related to these concepts.

The third and presumably most important issue is that the operationalization of meaning in this field of research has so far been restricted to only some selective aspects of sensorimotor information (e.g., shape, movement). The role of affect as a most basic human experience shaping the learning, representation, and processing of language (Jacobs, Hofmann, et al., 2016; Kousta et al., 2011, 2009; Schmidtke & Aryani, 2015; Vigliocco et al., 2009) has been surprisingly neglected. Indeed, affective dimensions of words, in particular, valence and arousal, are essential features defining a two-dimensional semantic space allowing for a very basic and potentially the most relevant distinction between different concepts; as empirically established by semantic differential (Osgood, 1952). In an attempt to provide a quantitative measure for words' meaning, Osgood (1952) defined 100 different lexical dimensions and asked participants to allocate the meaning of words for each dimension in an experiential continuum definable by a pair of polar terms (e.g., soft/hard, long/short, angular/rounded). Factor analyses conducted on the wide variety of verbal judgments indicated that most of the variance was accounted for by three major semantic dimensions: The two primary dimensions of 'valence' and 'arousal', and a third, less strongly-related dimension (in terms of the explained variance) of 'dominance' or 'control' (Mehrabian, 1980). Therefore, these factors have been considered basic dimensions of the semantic space within which the meaning of any concept can be specified.

Moreover, the expression and perception of affective states are fundamental aspects of human communication (Darwin, 1888; Wundt, 1908) that have been proposed as the original impetus for language evolution; with mimetic vocalization of emotional sounds supposedly allowing early hominids to efficiently share biologically significant information (Darwin, 1888; Ma & Thompson, 2015; Panksepp, 2010). Therefore, we would expect the effect of iconicity to be most evident in the communication of affect and in the relationship between words' affective sound (i.e., how emotionally words sound) and words' affective meaning (i.e., their position in the bi-dimensional affective space of lexical valence and arousal). Thus iconicity can serve as an interface for accomplishing the need to map linguistic form to human affective experience as a vital part of meaning making.

6.2.2 An embodied view on affective meaning

It is important to consider that the notion of "affective meaning" may not be shared by all theories on linguistic meaning. Our approach in this work is based on an embodied view of language which proposes that meaning is grounded in behavior (perception and action) and neural circuitry of the producer or the interpreter of linguistic signs (Bühler, 1934; Gallese & Lakoff, 2005; Glenberg, 2010; Jacobs et al., 2015; Jacobs, Hofmann, et al., 2016; Meteyard, Cuadrado, Bahrami, & Vigliocco, 2012; Vigliocco et al., 2009). Ultimately, part of the meaning of any utterance is its effect on the (physical and emotional) well-being of the person saying or hearing it, and everything that matters is represented in each individual person's brain and its neurophysiological systems. Presumably, the most fundamental such system is affect: in order to make meaning, we need to know what object/event in our environment requires us to react with alert or to keep calm, to approach or to withdraw. Moreover, the ability to distinguish between such affective contexts or reactions is linked to attention systems that select specific sensory input for further processing, and also to motor systems that select specific actions for output. Both systems (i.e., sensory and motor) provide crucial information for the construction of meaning by language users. Findings on the role of affective meaning in modulating various cognitive processes, such as learning, memory, attention or language processing, (Glenberg, Webster, Mouilso, Havas, & Lindeman, 2009; Jacobs, Hofmann, et al., 2016; Kousta et al., 2009; Vigliocco et al., 2009) support the idea that affective meaning is intertwined with other lexico-semantic aspects and has an essential and basic contribution to the process of meaning making.

6.2.3 The present study

We addressed the above-mentioned problems apparent in previous research by focusing on the affective meaning of real words and investigated whether participants were implicitly influenced by words' sound while giving a rating on emotion expressed in words' meaning.

Specifically, we aimed to challenge the established notion that assigning (affective) meaning to words is merely determined by words' semantic content and by an associative and per se arbitrary relationship between the signifier (sound image) and the signified (concept) – as encouraged by a leading principle of modern linguistics (Saussure, 1916/2011). Instead, we propose that the overall affective meaning of a word is, in addition to the word's semantic content, co-determined by inherent qualities of the signifier and by the percept derived from words' acoustic-phonetic features (i.e., the *affective sound*). Note that our use of the term affective sound in this paper refers exclusively to phonological constituents of words themselves and not to speaker-related issues such as intonation or how a word is spoken.

Our main hypothesis is motivated by research on nonverbal emotional vocalization and, in particular, emotional prosody which has shown that the emotional significance of a sound can be detected, and hence be integrated with higher-order cognition, even when the attentional focus is not directly on the emotional cues of the sound (Brück, Kreifelts, & Wildgruber, 2011; Frühholz, Trost, & Kotz, 2016a; Grandjean et al., 2005; Schirmer & Kotz, 2006). Such emotional cues have been shown to be engaged even in silent reading by means of cross-sensory input from the visual cortex into the auditory cortex and affective regions in the brain (Brück, Kreifelts, Gößling-Arnold, Wertheimer, & Wildgruber, 2014; Perrone-Bertolotti et al., 2012); as put forward by theories of embodied cognition and perceptual simulation. On the other hand, phonemes and their combinations (as in words) are characterized by a number of acoustic features that overlap with those that modulate emotional vocalization and emotional prosody (e.g., sound formants, sound intensity). Therefore, the specific sound profile of any word in a language can theoretically be attributed to a specific emotion as perceived by the listener. We thus hypothesized that the process of affective lexical evaluation—as higher order cognition—is influenced by words' phonology: that is, their phonologically recoded neuronal representation of the acoustic features corresponding to phonological word forms (Braun et al., 2009; Breen, 2014; Mesgarani, Cheung, Johnson, & Chang, 2014; Ziegler & Jacobs, 1995).

We used a large-scale normative database including rating values for affective meaning of words that were evaluated by at least 20 subjects/item. In line with Osgood's semantic differential (Osgood, 1952), such databases usually contain two types of ratings. The first component concerns emotional valence going from unpleasant to pleasant. For instance, words such as "murder", "poison", and "virus" are commonly rated as "very unpleasant" whereas "freedom", "love", and "life" lie on the other extreme end. The second type of ratings addresses the degree of emotional arousal ranging from excited (e.g., “nightmare”, “sex”, and “courage”), to calm (e.g., “health”, “massage”, and “peace”). Using these ratings as a measure of words’ affective meaning, we tested the null hypothesis (H0) that explicit evaluations of affective meaning solely reflect written words’ semantic content, against the alternative hypothesis (H1) that phonological word forms also contribute to valence and arousal ratings so that a statistically significant portion of their variance can be accounted for by words’ acoustic features. For instance, harsh-sounding words might make people feel more aroused so that they implicitly give a higher arousal rating, even though they are instructed to only focus on the lexico-semantic aspect of words.

In order to test H1, we chose a computational approach that employs signal averaging to amplify the potential effect of sound on meaning. Subsequently, we quantified a *Phonological Affective Potential* (PAP) of words, separately for arousal (PAP_{aro}) and for valence (PAP_{val}), both estimates representing the influence of words’ affective sound on their affective meaning. The goal of this work is to examine the psychological reality of the PAP, and to test whether the PAP is linked to (and derived from) words’ acoustic features that we extracted from their spoken forms after synthesizing them (study 1). Note that according to H0 the amount of variance in the PAP that depends on words’ acoustic features should not be greater than chance level and therefore statistically not significant. Next, we tested the association of the PAP with words’ affective sound assessed in two behavioral studies using different methods. We first asked participants to rate the affective sound of printed words while suppressing words’ meaning (study 2a), and second, we employed auditory presented pseudowords and acoustic models to predict words’ affective sound based on their acoustic features (study 2b). We then compared the PAP with these two independent measures of affective sound to test for their potential associations. Finally, in order to identify the perceptual cues potentially underlying the effect of implicit sound on meaning, we separately tested the relationship between words’ acoustic features and words’ PAP and the two independent measures for words’ affective sound.

Although previous studies provided first affirmative support for the affective potential of some phonological units (Aryani, Jacobs, & Conrad, 2013; Aryani, Kraxenberger, Ullrich, Jacobs, & Conrad, 2016; Myers-Schulz, Pujara, Wolf, & Koenigs, 2013; John J Ohala, 1994; Ullrich, Kotz, Schmidtke, Aryani, & Conrad, 2016), ours is the first study demonstrating the psychological reality of phonemes' affective potential and the contribution of words' implicit sound to meaning making for real words and across a language's lexicon.

6.3 Quantifying the Phonological Affective Potential

6.3.1 Material

The Berlin Affective Word List (Võ et al., 2009) (BAWL) was used as a normative database containing a representative mass of 2694 German words that has been cross-validated in various empirical studies regarding experiential, behavioral, and neurobiological levels of analysis (Jacobs et al., 2015). The BAWL includes words from different classes (nouns, verbs, and adjectives) that were selected based on the following main criteria: to include a) the most frequently used German words, b) as many words as possible with an apparent relation to affect regardless of whether this would result in more or less extreme values of valence and arousal, and c) a critical mass of theoretically neutral words. As a consequence, the BAWL contains a relatively elevated percentage of emotion-laden words. However, valence and arousal values of words in the BAWL are spread across the entire range of both valence and arousal in order to make it an optimal tool for selecting verbal material for all kinds of research questions on affective language processing. Henceforth, we refer to these valence and arousal values representing words' affective meaning as *Affective Meaning Ratings*.

6.3.2 Method

We operationalized H1 by assigning two statistical components to each of the *Affective Meaning Ratings* in the BAWL (separately for arousal and valence): for a word composed of phonemes W_i [$ph_1, ph_2..., ph_n$], the rating value was considered to reflect a first component stemming from an explicit evaluation of the word's *Affective*

Semantic Content, and a second phonological component, the PAP (*Phonological Affective Potential*), which reflects the contribution of the affective potential of phonemes to the total rating value. Assuming a simple additive model, the rating for a given word in the database can thus be modelled as the sum of these two components plus an error term ϵ

$$\textit{Affective Meaning Rating}\ (W_i) = \textit{Affective Semantic Content}\ (W_i) + \mathrm{PAP}_i\ [ph_1, ph_2, \ldots, ph_n] + \varepsilon \quad (1)$$

We estimated the PAP by averaging the potential affective effects of all phonemes regardless of their position in the word. The PAP thus can be hypothesized as a function of the *Phonological Affective Value* (PAV) of each phoneme: that is, the contribution of the affective quality of each single phoneme to the total rating value for each word:

$$\mathrm{PAP}_i\ [ph_1, ph_2, \ldots, ph_n] = \text{Mean}\ [\mathrm{PAV}\ (ph_1), \mathrm{PAV}\ (ph_2), \ldots, \mathrm{PAV}\ (ph_n)] \qquad (2)$$

In order to quantify the PAV(ph_j), we considered it to be a signal masked by 'noise': that is, the *Affective Semantic Content* (W_i) and the error term in Eq. 1. We thus attempted to minimize the effect of noise while increasing the signal-to-noise-ratio (SNR) by averaging. That is, for each phoneme, we calculated the average rating values of words that contain this phoneme (Figure 6.1-A). This way, the *Affective Semantic Content* (W_i) and the error term ϵ nearly cancel out, and the average ratings are approximately associated with the potential contribution of each single phoneme (i.e., the PAV) to the rating value.

6.3.3 Corpus Preparation:

In order to have an adequate number of repetitions and hence improve the SNR, we chose only those phonemes with a frequency of appearance higher than 30 in the database (mean frequency = 322). This led to the exclusion of 120 words that contained phonemes with a lower frequency, including those that are not a part of German phonology. Overall 12 phonemes were excluded, seven with a frequency of seven to 30 (ʒ, ɛː, |:, j, pf, tʃ, /) and five which did not belong to German phonology.

6.3.4 Calculation of PAPs

For a given word W_k composed of n phonemes $\{Ph_1, Ph_2, \ldots, Ph_n\}$ with a rating R_k (valence or arousal), we first defined a membership function α as follows:

$$\begin{Bmatrix} Ph_i \in W_k \Rightarrow \alpha(Ph_i, W_k) = 1 \\ Ph_i \notin W_k \Rightarrow \alpha(Ph_i, W_k) = 0 \end{Bmatrix}$$

For each of the 36 phonemes in the database (Ph_i) the *PAV* was calculated (separately for arousal and valence) as follows:

$$PAV(Ph_i) = \frac{\sum_{j=1}^{N} \alpha(Ph_i, W_j) \times R_j}{\sum_{j=1}^{N} \alpha(Ph_i, W_j)}$$

Where W_j is the j^{th} word in the database, and N is the number of whole words (= 2574). Results are shown in S1 (supplementary data).

For each word in the database (W_k), we then calculated its PAP (again, separately for arousal and valence) by averaging across all PAVs (see Figure 6.1-A). The *Phonological Affective Potential* (for valence or arousal) for a given word W_k will then be:

$$PAP(W_k) = \frac{\sum_{i=1}^{n} \alpha(Ph_i, W_k) \times \frac{\sum_{j=1}^{N} \alpha(Ph_i, W_j) \times R_j}{\sum_{j=1}^{N} \alpha(Ph_i, W_j)}}{\sum_{i=1}^{n} \alpha(Ph_i, W_k)}$$

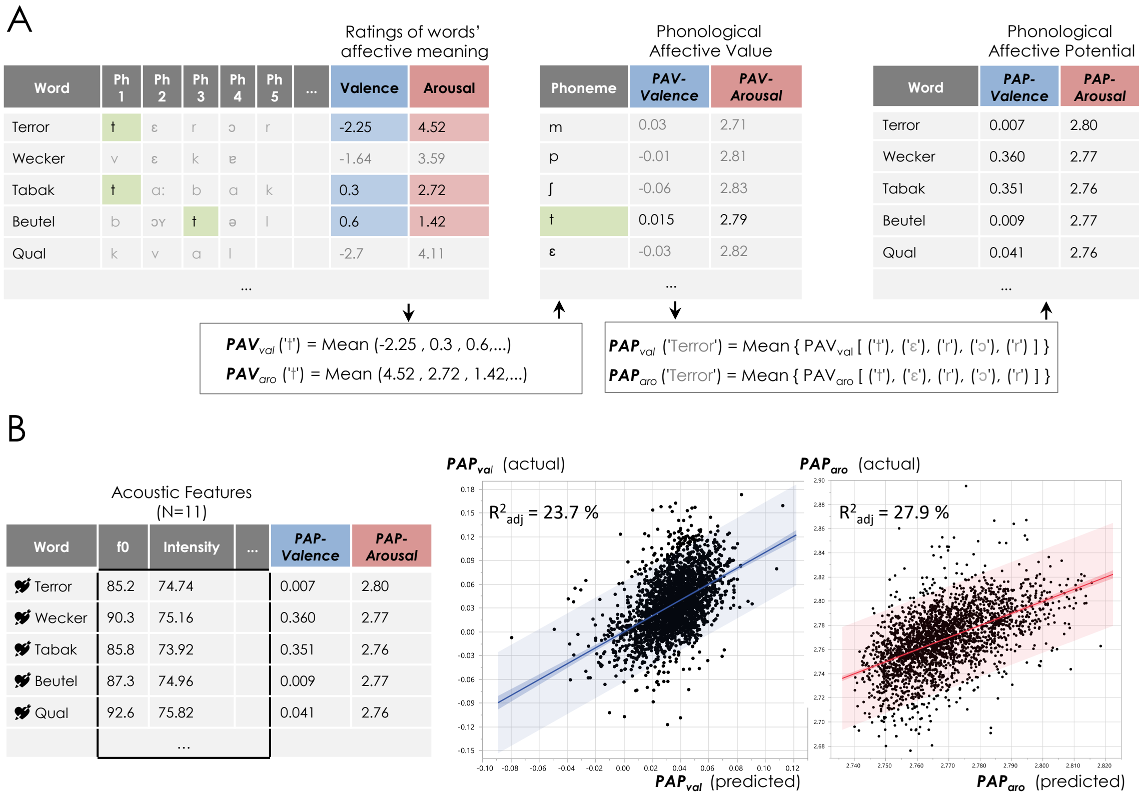


Figure 6.1. **A)** Words in the normative database (BAWL) were segmented and coded for the presence or absence of a given phoneme (here exemplified by the phoneme /t/). The phonemes were analyzed one-by-one to determine their potential effect on valence and arousal ratings. The potential affective effect caused by each single phoneme (i.e., PAV) was computed as the average of valence or arousal ratings of words containing this specific phoneme. The PAP of each word was calculated as the average of all its PAVs. **B)** Words were synthesized and their extracted acoustic features were used in two multiple linear regression models as predictors for the PAP of arousal (right) and valence (left). The acoustic variables (11 in total) accounted for 27.9% and 23.7% of the variance in PAP_{aro} and PAP_{val} respectively (study 1).

6.4 Study 1: Relating Words' Acoustic Features and PAP

If the PAP of written words is somehow linked to emotional cues present in their phonological forms, we would expect it to be related to the acoustic features of that form.

6.4.1 Method

To test the above relationship, we synthesized the words and extracted their acoustic features, focusing on a total of 11 features that are known to modulate emotional vocalization: fundamental frequency (f0; mean), sound intensity (mean and standard deviation), spectral center of gravity (mean), standard deviation of the spectrum, and sound formants (F1, F2, F3; means and bandwidths) (Brück et al., 2011; Juslin & Laukka, 2003; Sauter, Eisner, Calder, & Scott, 2010; Scherer, 2003).

It is worth pointing out that we deliberately opted for synthesizing the words rather than using a human speaker in order to prevent any undesired emotional prosody that might result from words' affective meaning: Human speakers tend to pronounce words with a prosodic intonation—independently from phonological content—consistent with words' meaning. By synthesizing the words, we distinctly separated our dependent and independent variables: PAPs (PAP_{aro} and PAP_{val}) on the one side, and acoustic features on the other side. Although the artificial nature of a synthesized voice could diminish the goodness of acoustic models, a positive result would all the more support the effects in question.

6.4.2 Synthesizing and Acoustic Analyses

We synthesized the words using the eSpeak (Duddington, 2008) as front-end to the male voice de4 from MBROLA (Dutoit, Pagel, Pierret, Bataille, & der Vrecken, 1996) which consists of a speech synthesizer; based on the concatenation of diphones, and of diphone databases. We abstained from the use of larger synthesis units (such as whole words or phrases as used in Variable Unit Concatenation systems) to avoid the potential effect of words' affective content on speakers' prosody as discussed above (van Heuven & Pols, 1993). Words were synthesized in a fixed carrier sentence, *Das Wort … wird oft verwendet* ("The word … is often used"). The rate of speech was set at 120wpm (words per minute). All spoken words were checked for intelligibility by two male native speakers (not otherwise involved in the study). Importantly, the speakers were not provided with the word list so that they had no expectations about the words' identity (van Bezooijen & van Heuven, 1997). Both speakers agreed on the intelligibility of all words: speaker1 marked four words and speaker2 marked seven words as poorly synthesized; however, they found all words still intelligible. We extracted the acoustic features of words using the speech analysis software PRAAT (Boersma & Weenik, 1996). We extracted the mean of fundamental frequency f0 (time step=.01, min=75Hz, max=300Hz), the mean and standard deviation of intensity (time step=.01), and the mean and the bandwidth of the first three formants F1, F2 and F3 (time step=.01) from the spectral representation of the sound. Finally, the spectral centroid (spectral center of gravity) and the standard deviation of the spectrum were computed on the basis of fast Fourier transformations (time step=.01, min pitch= 75Hz, max pitch= 300Hz).

6.4.3 Results and Discussion

We computed two multiple linear regression models to predict the PAP_{aro} and the PAP_{val} using the above distinctive acoustic features as regressors (N = 11). The acoustic analyses reported next were carried out on all of the 2574 words in the database. The results are summarized in Figure 6.1-B and S2. For both arousal and valence, PAPs were significantly predicted by the distinctive acoustic variables (both Ps < 0.0001), the variance accounted for being 27.9% for the PAP_{aro} and 23.7% for the PAP_{val} (both R^2 adjusted).

Words in the database that are derived from the same stem or root morpheme (e.g., '*terror*' and '*terrorize*') are likely to have both similar phonological structure and similar semantic content. This could potentially bias the relationship between PAP and the way words sound. Hence, to ensure unbiased estimation, we selected all monosyllabic words from the database (N = 289) and repeated the above analysis steps (including new calculations of PAPs) for this subset comprising only monomorphemic words for which stem repetition was precluded. This time, the 11 acoustic variables accounted for 29.3% of the variance in PAP_{aro} and 26.6% in PAP_{val} (both R^2 adjusted, Ps < 0.0001). The successful outcome and even larger portion of explained variance corroborates our previous results and validates the method used to uncover the effects of phonological units. We expected to obtain a better approximation for PAV (and consequently PAP) when the number of phonemes in a word is reduced, as is the case for monosyllabic words.

By showing that a considerable part of the variance in PAPs depends on the acoustic features of the spoken word forms, we could reject H0 stating that PAPs are a mere product of chance. Instead, H1 was supported: acoustic features of phonemes significantly co-determined words' affective ratings even when they are visually presented (and silently read). This suggests that the contribution of phonological units to the ratings of words' affective meaning—as reflected in the PAPs—emerges from a representation of acoustic properties of words in spoken form. We take this as a first support for the validity or psychological reality of the effect in question.

Having shown that PAPs of written words are significantly associated with the acoustic profile of their spoken form, we next asked whether this association is based on the words' affective sound as assessed in two independent ways (study 2a and 2b).

The following studies were approved by the ethics committee of the Freie Universität Berlin and were conducted in compliance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). All participants gave their consent (in written form for study 2a, and online for study 2b) prior to participating in the study.

6.5 Study 2a: Measuring Words' Affective Sound via Rating

6.5.1 Stimuli

The stimuli were the 2574 words from the BAWL used in the previous analyses.

6.5.2 Participants

A total of 272 participants were recruited by flyers, email contacts, and Facebook posts, who then rated the words either for valence or for arousal. Of these, 135 participants (82 females, age=21.3 ± 4.6) rated exclusively for arousal and 137 participants (92 females, age=23.6 ± 2.9) for valence. Participants were mostly students from the Freie Universität Berlin who received either psychology course credit or 5 Euros for their participation. All participants reported normal or corrected-to-normal vision and were native German speakers.

6.5.3 Procedure

Words were presented visually. A very similar set of instructions to those used to rate the words' affective meanings (Võ et al., 2009) was applied here, with one minor modification. Participants were instructed that they would have to suppress the meaning of words and only pay attention to their sound, and this was repeatedly emphasized through the instruction process. We also incorporated the self-assessment manikins (SAM) that were used in the ANEW study (Bradley & Lang, 1999). Words were randomly divided into 8 different lists each of which included about 335 items. Words were then rated on both affective sound of valence and affective sound of arousal by different groups of participants in order to exclude the possibility of mutual influence between valence and arousal ratings. The affective sound of valence was rated on a 5-point scale ranging from -2 (*sehr negativ* / "very negative") through 0 (*neutral* / "neutral") to +2 (*sehr positiv* / "very positive"). The 5-point affective sound of arousal scale ranged from 1 (*sehr beruhigend* / "very calming") to 5 (*sehr aufregend* / "very exciting"). The items were randomly presented to minimize primacy or recency effects. On average, the tasks were completed in approximately 25 minutes.

6.5.4 Analysis

Each word was rated by an average of 19.7 participants for valence and 20.4 participants for arousal. In order to assess the degree of agreement among raters, the Interclass Correlation (ICC) was computed for both arousal and valence ratings. Results showed a higher value for arousal (ICC=0.43) than for valence (ICC=0.31), indicating a rather poor degree of agreement.

Even though participants were asked to only focus on the affective *sound* of words, their ratings were likely "contaminated" by words' semantic content, since semantic representations are automatically activated during word recognition. This was evident in the correlations between our ratings of affective sound and the original *Affective Meaning Ratings*: r = 0.32 for arousal and r = 0.22 for valence. To eliminate the undesired effect of words' *Affective Semantic Content* from our ratings of affective sound, we opted for the most conservative approach. We first regressed the PAPs on the *Affective Meaning Ratings*—separately for arousal and valence—and used the residuals as a statistical estimate for words' *Affective Semantic Content* (cf. Eq. 1). In a next step, we regressed the estimate for words' *Affective Semantic Content* on our ratings of affective sound and used the z-transformed residuals of this regression as independent measures of words' affective sound. This way, the potential effect of *Affective Semantic Content* was partialed out of rating values of affective sound. The substantially weaker correlations between these "decontaminated" residuals and the original *Affective Meaning Ratings* validated our method: r = 0.1 for arousal, and r = 0.04 for valence. These two decontaminated residuals (for arousal and valence) were then used as estimates of the words' affective sound.

In the following, we refer to these two measures as *Affective Sound Ratings*; in short: AS-R_{aro} (for arousal) and AS-R_{val} (for valence).

6.5.5 Results and Discussion

The correlations between AS-R_{aro} and AS-R_{val} on the one hand, and PAP_{aro} and PAP_{val}, on the other hand, were highly significant: r = 0.5, for arousal, and r = 0.25, for valence (both Ps < 0.0001). A similar analysis was performed for the subset of monosyllabic words (N=289): AS-R_{aro} and AS-R_{val} were also significantly correlated with the corresponding PAP_{aro}: r = 0.46, and PAP_{val}: r = 0.32, respectively (both Ps < 0.0001).

These results indicate that the contribution of phonological units to words' affective meaning ratings (i.e., PAPs) is associated with the affective sound of these words, thus providing further support for the psychological reality and the affective nature of PAPs.

6.6 Study 2b: Predicting Words' Affective Sound via Acoustic Models

In the previous study, the poor ICC values suggested that subjective judgments about the affective sound of a word while trying to suppress its meaning can be a difficult task. In this study, we therefore aimed to provide a new measure of affective sound by using meaningless pseudowords that would allow participants to better focus on the sound. We therefore generated and presented pseudowords in auditory form and collected ratings of their affective sound. By extracting the acoustic features of pseudowords and using them as predictors we developed acoustic models capable of predicting the variation in the ratings. Such independent models can then be applied to any word-like item in auditory form to predict its affective sound solely based on its acoustic features, including the real words from the previous studies. Note that since the pseudowords had to be presented to and rated by human subjects, for this task—unlike in Study1—we used a human voice rather than a synthesizer to generate naturally sounding stimuli and to prevent potential distortion effects of sound peculiarity. However, as pseudowords lack semantic content there was no concern about the influence of meaning on emotional prosody as in the study1.

6.6.1 Stimuli

To generate pseudowords representative for the phonotactics of German, we used the Wuggy algorithm (Keuleers & Brysbaert, 2010) which generates pseudowords that match a given word template in sub-syllabic structure and transition frequencies, thus obeying a language's phonotactic constraints. Since the pseudowords had to be spoken and rated, to avoid obscureness we restricted the list of word templates to those having up to three syllables and 10 letters. We then chose the first 1500 most frequent nouns from CELEX (Baayen, Piepenbrock, & van H, 1993). For each word, we adapted the program to generate five pseudoword alternatives using Wuggy's default setting. Candidate pseudowords which differed in fewer than two letters (whether added, deleted or substituted) from the nearest real word were excluded due to their similarity to real words (Coltheart distance = 1). For words with more than one remaining pseudoword alternative, the one with a highest Levenshtein distance (Yarkoni et al., 2008) was selected. The list of pseudowords was checked for pseudohomophones and a too high similarity to real words. Thus, 187 items were excluded: for example, *beim* (similar to the short form of "*bei dem*"= for something), *absads* (similar to the word "*Absatz*"=paragraph). In addition, because of phonotactic problems mostly caused by illegal or very rare grapheme combinations 190 items were excluded: for example Weckbeveuz, Ymiön, by two native speakers. The remaining 1123 pseudowords were selected for recording.

A professional male actor was recruited in Berlin, Germany, who was a native speaker of German. He had graduated from professional acting school and was regularly employed in radio, television, and stage work. He was paid to participate. Pseudowords were spoken in a list-like manner to prevent affective prosody and were recorded in the "Leibniz-Zentrum Allgemeine Sprachwissenschaft" in Berlin in a professional sound recording booth using a "Sennheiser MKH20" microphone and

"Ultra Gain MIC-2000" preamplifier. The audio signal was recorded using the DAT-recorder "TASCAM DA20MKII" with a sampling frequency of 48 kHz and 16 bits per sample.

6.6.2 Participants

A total of 169 participants were recruited by flyers, email contacts, and Facebook posts, who rated the pseudowords either for valence or for arousal. Of these, 85 participants (52 females, age = 26.7 ± 4.3) rated exclusively for arousal and 84 participants (42 females, age = 27.1 ± 3.8) for valence. Participants had the chance to win one of 10 Amazon coupons which were assigned randomly at the end of the study. All participants were native German speakers.

6.6.3 Procedure

In order to afford a convenient method of sampling that was more representative of the general population, the study was conducted online using the SoSci panel (Leiner, 2014). Adapting the instructions used for the original BAWL ratings for written words, participants were invited to carefully listen to the presented item and evaluate how positive or negative (in the case of valence) and how exciting or calming (in the case of arousal) the pseudowords sounded. During the rating process, a "replay" button was offered to provide participants with the opportunity of repeated listening to each presented item. We also incorporated the self-assessment manikins (SAM) that were used in the ANEW study (Bradley & Lang, 1999). Importantly, participants were instructed to give their ratings solely based on the sound aspect of items and not based on their similarity to real words. In order to prevent participants from giving a rating for a similar sounding word, a button labeled "concrete word" was provided next to the rating scale, and participants were instructed to use it in case an item might remind them of a German word. 28 items labelled as "concrete word" by more than 50% of participants were then excluded, leaving 1095 items for further analysis. Pseudowords were randomly divided into 4 different lists, each including about 280 pseudowords. The order of presentation was pseudorandomized for each participant. On average, the task was completed in approximately 15 minutes.

6.6.4 Analysis

Affective ratings were obtained for 1095 pseudowords with 17 ratings per item on average (17.2 for arousal, and 17.5 for valence). We extracted the 11 acoustic features from the spoken pseudowords (see study 1) and performed two multiple regression models using them as predictors of the ratings separately for arousal and valence. These features accounted for 56.3% of the variance in arousal ratings and 11.2% for valence (both R^2 adjusted, Ps < 0.0001, Figure 6.2).

Since our ultimate goal was to predict the affective sound of real words, in order to assess how the results of the above models generalize to an independent data set (i.e., real words), we used two-fold cross-validation. The dataset was randomly shuffled into two subsets with equal size one for training and one as a test set, and vice versa. Model accuracy for each run was 57.3% and 52.6% (both R^2 adjusted, Ps < 0.0001) for the arousal model, and 10.1% and 9.9% for the valence model (both R^2 adjusted, Ps < 0.0001). These are very robust results in terms of explained variance compared to the original models.

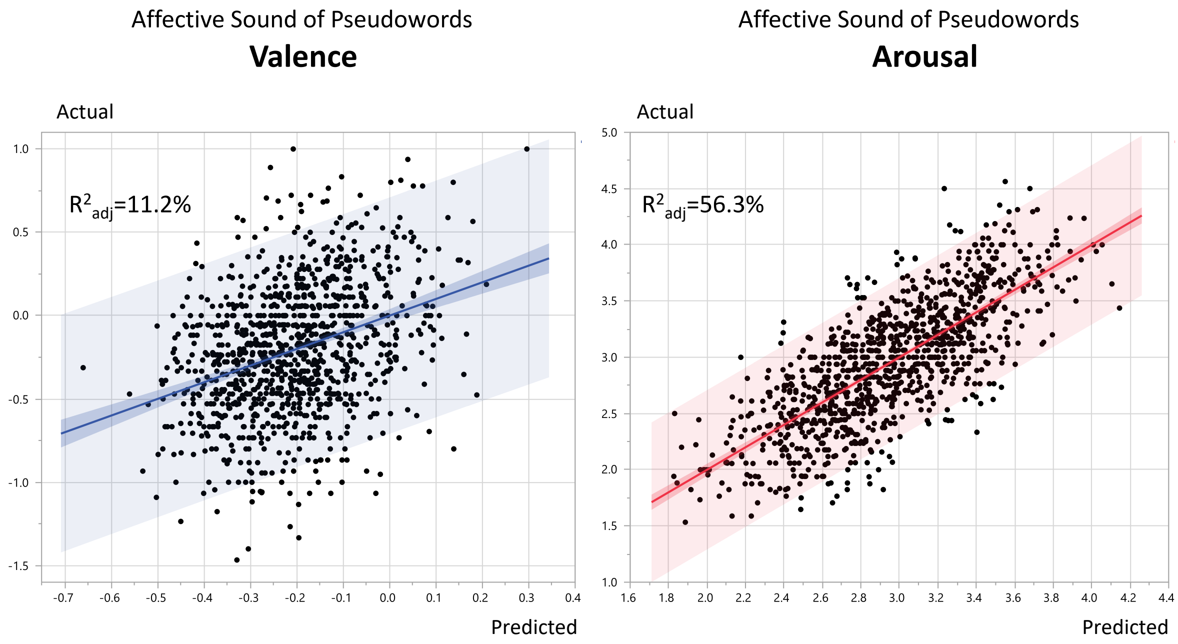


Figure 6.2. Acoustic features of pseudowords (N=11) significantly predicted the ratings of their affective sound: 11.2% for valence (left) and 56.3% for arousal (right).

6.6.5 Results and Discussion

The degree of agreement among raters, compared to the results of study 2a, was considerably higher for both valence (ICC=0.61) and arousal (ICC=0.86). The substantial amount of variance accounted for in our regression model for arousal indicates that the affective sound of word-like stimuli could be mapped out in terms of their acoustic cues; a strong evidence for acoustic features to possess affective value on their own. A closer look at the variation in ratings revealed a smaller relative standard deviation for valence (13%) than for arousal (18%), suggesting a lower consensus among participants when rating valence. The considerably higher degree of explained variance for arousal as compared to valence supports the idea that speech sounds primarily signals the sender's arousal state (and their valence state only to a smaller degree) (J. A. Bachorowski & Owren, 1995; Sauter et al., 2010); we will discuss this finding more fully later in this article.

We next took the two acoustic models (i.e., the linear equations in S3 and S4) resulting from the pseudoword data and applied them to the extracted acoustic features of the words in the database to predict words' affective sound. We refer to these predicted values for words' affective sound as *Affective Sound Predicted*; in short: AS-P_{aro} (for arousal) and AS-P_{val} (for valence).

The obtained predicted values for words' affective sound (i.e., AS-Ps) were then compared with PAP: The AS-P_{aro} and the AS-P_{val} of words were significantly correlated with the PAP_{aro}: $r = 0.47$, and with the PAP_{val}: $r = 0.36$, respectively (both Ps < 0.0001). Again, similar results were obtained for monosyllabic words (N=289): AS-P_{aro} was significantly correlated with PAP_{aro}: $r = 0.45$, $P < 0.0001$, and AS-P_{val} was significantly correlated with PAP_{val}: $r = 0.42$, $P < 0.0001$.

These significant associations between words' PAP and their affective sound of words—independently predicted from acoustic features—add additional support for our H1. In addition to the direct correlation between PAPs and words' affective sound, as captured by AS-Ps (AS-P_{aro} and AS-P_{val}), we tested the relationship between AS-Ps and those proportions of variance in PAPs that we could account for by means of acoustic features in the first analysis: that is, the predicted values for the PAP_{aro} and the PAP_{val} in the first multiple regression models (Study1, Figure 6.1-B) that were calculated with the same acoustic variables as regressors. Results showed high correlations between AS-P_{aro} and the predicted values for PAP_{aro}: $r = 0.88$, and between AS-P_{val} and the predicted values for PAP_{val}: $r = 0.71$ (both Ps < 0.0001). This suggests that the PAPs are based on the same distinctive acoustic features that participants used to evaluate the affective sound of pseudowords, thus, again, providing strong evidence for the association between PAPs and affective sound, and that a significant portion of variance in the ratings of words' affective meaning is due to how words affectively sound.

Furthermore, we tested the reliability of our two different measures of words' affective sound as described in Study 2a and Study 2b, to investigate their consistency in capturing the same concept. For this, we compared the values resulting from these completely independent methods for measuring words' affective sound. Results showed significant correlations between the measure of affective sound based on the direct rating value (i.e., AS-R, Study 2a) and the predicted values of affective sound based on acoustic features (i.e., AS-P, Study 2b) for both arousal: $r = 0.56$, $P < 0.0001$, and valence: $r = 0.49$, $P < 0.0001$. These results, together with the fact that PAPs are associated with words' affective sound, provide firm support for our H1 stating that the affective meaning of words is shaped by both words' semantic content and (implicit) affective sound.

6.7 Analysis of Words' Acoustic Profiles

Having shown a robust association between PAPs and two independent measures of affective sound (AS-R and AS-P), we continued with a more fine-grained analysis and asked whether the underlying acoustic features shaping PAP and both AS-R and AS-P, do so in identical or differential ways for these different measures. Thus, we examined the direct relationships between each acoustic feature and the PAP on the one hand, and our two measures of affective sound (i.e., the AS-R from study 2a and the AS-P from study 2b) on the other.

6.7.1 Method

We constructed acoustic profiles based on the strength and direction of correlations between each of 11 acoustic variables with PAP, AS-R, and AS-P (see Figure 6.3, see S5 for correlation coefficients).

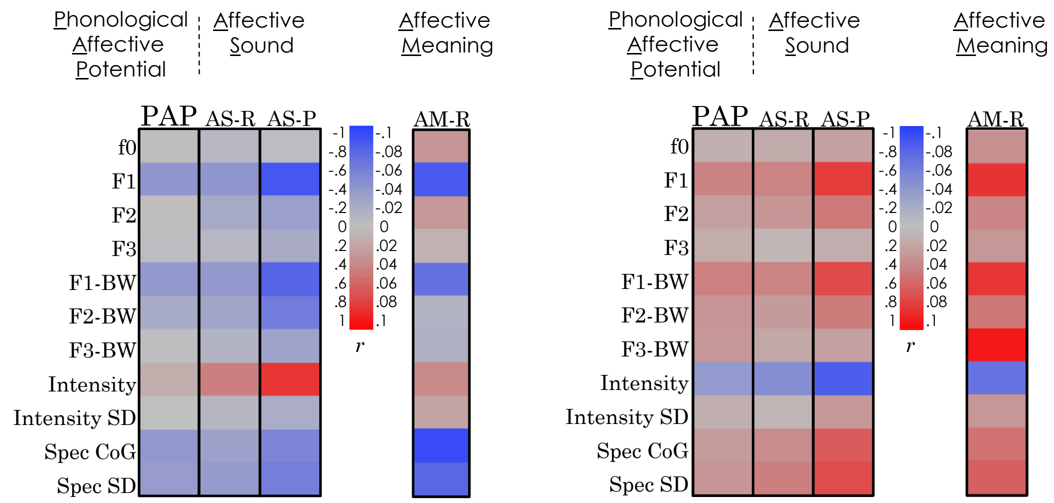
6.7.2 Results and Discussion

For the arousal dimension, all single correlations (N=3x11) were highly significant ($Ps < 0.0001$). Notably, the correlations between each acoustic variable and PAP_{aro} were always in the same direction as correlations between this specific acoustic variable and both measures of affective sound for arousal (i.e., the $AS\text{-}R_{aro}$, and the $AS\text{-}P_{aro}$) resulting in highly similar acoustic profiles for all three measures (Binomial test: $X \sim B\ (11,\ 0.5)$, $p\ (X \geq 11) = 0.0005$). A similar pattern was observed for valence. The PAP_{val} was significantly correlated with seven acoustic variables, and, importantly, these correlations were, again, always in the same direction as for the acoustic variables and both measures of affective sound for valence (i.e., the $AS\text{-}R_{val}$, and the $AS\text{-}P_{val}$) again resulting in highly similar acoustic profiles for all of three measures (Binomial test: $X \sim B\ (7,\ 0.5)$, $p\ (X \geq 7) = 0.007$).

All correlations in the acoustic profile of arousal remained significant after Bonferroni correction for multiple comparisons. For the acoustic profile of valence, however, the correlation between the third formant (F3) and PAP_{val} did no longer reach statistical significance after Bonferroni correction. But, still, acoustic profiles for all of three measures ($AS\text{-}R_{val}$, $AS\text{-}P_{val}$, PAP_{val}) remained highly similar (Binomial test: $X \sim B\ (6,\ 0.5)$, $p\ (X \geq 6) = 0.015$).

These results go beyond the simple relationships between the PAPs and the affective sound of words (as captured by AS-P and AS-R); moreover, they show that the acoustic features that underlie PAPs contribute in very similar ways to the perception of words' affective sound. We interpret this as strong support for PAPs being determined by affective perceptual cues within phonological word forms.

A



B

Valence

	PAP$_{val}$	AS-R$_{val}$	AS-P$_{val}$	AM-R$_{val}$
f0	0.300	<0.0001	0.0398 [BF]	0.220
F1	<0.0001	<0.0001	<0.0001	0.0008
F2	0.179	<0.0001	<0.0001	0.233
F3	0.049[BF]	<0.0001	<0.0001	0.634
F1-BW	<0.0001	<0.0001	<0.0001	0.0068[BF]
F2-BW	<0.0001	<0.0001	<0.0001	0.600
F3-BW	0.338	<0.0001	<0.0001	0.489
Intensity	<0.0001	<0.0001	<0.0001	0.127
Intensity SD	0.809	<0.0001	<0.0001	0.381
Spec CoG	<0.0001	<0.0001	<0.0001	0.0001
Spec SD	<0.0001	<0.0001	<0.0001	0.0027

Arousal

	PAP$_{aro}$	AS-R$_{aro}$	AS-P$_{aro}$	AM-R$_{aro}$
f0	<0.0001	<0.0001	<0.0001	0.086
F1	<0.0001	<0.0001	<0.0001	<0.0001
F2	<0.0001	<0.0001	<0.0001	0.0437[BF]
F3	<0.0001	<0.0001	<0.0001	0.134
F1-BW	<0.0001	<0.0001	<0.0001	<0.0001
F2-BW	<0.0001	<0.0001	<0.0001	0.0145[BF]
F3-BW	<0.0001	<0.0001	<0.0001	<0.0001
Intensity	<0.0001	<0.0001	<0.0001	0.0009
Intensity SD	<0.0001	<0.0001	<0.0001	0.139
Spec CoG	<0.0001	<0.0001	<0.0001	0.0095[BF]
Spec SD	<0.0001	<0.0001	<0.0001	0.0021

Figure 6.3. A) Acoustic profiles were constructed (using correlation cell plot) based on the strength and direction of correlations between the estimated effect of words' phonology on the evaluation of their affective meaning (i.e., Phonological Affective Potential: PAP), the two measures of words' affective sound (i.e., Affective Sound-Ratings: AS-R (study 2a), Affective Sound-Predicted: AS-P (study 2b), and ratings of words' affective meaning (i.e., Affective Meaning-Ratings: AM-R) on the one hand, and 11 acoustic variables on the other hand (left for valence, right for arousal). **B)** The correlation probabilities are shown in the table. Correlations not surviving Bonferroni correction for multiple comparisons are marked with "BF" (Bonferroni Failed).

Abbreviations: BW= Bandwidth, SD= standard deviation, Spec=Spectral, CoG = Centre of Gravity, r= correlation coefficient.

6.8 The Direct Effect of Sound on Words' Affective Meaning

Here, we asked whether the contribution of words' (implicit) sound to words' affective meaning can be directly observed at the level of original valence and arousal ratings in the database: that is, before estimating the effect through our statistical operationalization for the PAP. In other words, if an *Affective Meaning Rating* consists of *Affective Semantic Content* and *Phonological Affective Potential* (PAP), as formulated in equation 1, we would expect that the same acoustic features shaping the PAP should be reflected, though to a lesser degree, in *Affective Meaning Ratings*. That is the effect of words' acoustic features on words' affective meaning should be observable directly at the level of *Affective Meaning Ratings*.

6.8.1 Method

We constructed acoustic profiles for *Affective Meaning Ratings* by calculating correlations between *Affective Meaning Arousal-Ratings* (in short AM-R_{aro}) and *Affective Meaning Valence-Ratings* (in short AM-R_{val}) and each of the acoustic variables across all words in the database (N=2574).

6.8.2 Results and Discussion

From the total of 11 acoustic variables, eight variables in the acoustic profile for arousal (five variables after Bonferroni correction) and four in the acoustic profile for valence (three variables after Bonferroni correction) were significantly correlated with AM-R_{aro} and AM-R_{val}, respectively (Figure 6.3 and S5). Most importantly, those acoustic features that significantly correlated with AM-R_{aro} and AM-R_{val} always showed an association in the same direction as the one between the acoustic features and respective PAP, as well as both measures of words' affective sound (i.e., the AS-R and AS-P). Again, these results support the direct relationship between words' acoustic features and ratings of affective meaning.

Together with our previous findings, these data suggest that the process of meaning making is not solely determined by arbitrary mappings between words' phonology and concepts they refer to. Rather, words' acoustic profiles provide affective perceptual cues that language users implicitly use to construct words' affective meaning.

6.9 Acoustic Phonetic Cues Underlying the Effect of Sound on Meaning

Revealing the perceptual acoustic cues likely underlying the effect of implicit sound on affective meaning, we performed further analyses to explore phonetic features potentially causing this effect.

The consistently negative correlations between sound intensity and each of the four arousal-based measures: PAP_{aro}, $AS\text{-}R_{aro}$, $AS\text{-}P_{aro}$, and $AM\text{-}R_{aro}$ (Figure 6.3) deserves a more detailed discussion as arousal usually increases with sound intensity when the latter is experimentally manipulated. Note that all words and pseudowords were spoken with the same loudness, thus differences in sound intensity have to be tracked back to specific phonetic features of the words in the database.

6.9.1 Long vs. Short Vowels

A closer look at the spectrograms reveals that words with the highest sound intensity tend to include long vowels (e.g., Lohn /l oː n/ "*wage*", See /z eː/ "*lake*", see Figure 6.4-A). To systematically examine this potential relationship, we defined a Vowel Length Index as the average vowel length (short=1, long=2) over the word's syllables. This Vowel Length Index was significantly correlated with sound intensity across all words in the database: $r = 0.28$, $P < 0.0001$, suggesting a systematic relationship between the two measures. With regard to affective perception, note that long vowels are produced through a release of air from the mouth for an extended period of time which is a behavior similar to slow (vs. rapid) breathing that, in turn, is associated with decreasing (vs.

increasing) arousal (Boiten, 1998; Nyklíček, Thayer, & Van Doornen, 1997). This relationship between affective states and sound duration is also stressed in the motivation-structural rules hypothesis (Morton, 1977) stating that calls produced by mammals in aggressive circumstances, termed *barks* or *grunts*, are generally of shorter duration than those produced in appeasement contexts. On the other hand, at the spectrogram level, the sustained high amplitude for long vowels causes a larger integral of energy for the whole sound envelope leading to the negative correlation between arousal and sound intensity (see Figure 6.4A). Note also that the variation of intensity of sound over time (Intensity-SD) accordingly displays a positive correlation with arousal.

A comparison between the PAV of short vowels and their long counterparts (see S1) revealed the same pattern: each of the short vowels was perceived as "more arousing" than its long counterpart: PAV_{aro}(/a/) > PAV_{aro} (/aː/), PAV_{aro} (/ɔ/) > PAV_{aro} (/oː/), PAV_{aro} (/ʊ/) > PAV_{aro} (/uː/), PAV_{aro} (/ɪ/) > PAV_{aro} (/iː/), PAV_{aro} (/ɛ/) > PAV_{aro} (/ɛː/). A very similar pattern was revealed for valence values calculated for short and long vowels; with short vowels being more "negative" than their long counterparts – except for the short and long vowels /ɔ/ and /oː/, for which all calculated values were very close to zero.

In addition, short vowels tend to be followed by more consonants (i.e., more complex consonant clusters) than long vowels, and this complexity of subsequent consonant clusters may also hold partly responsible for the observed correlation between vowel length and arousal.

6.9.2 Voicing

Another phonetic feature directly related to sound intensity is 'voicing'. Voiced consonants are accompanied by vocal cord vibration that leads to an increase in sound energy compared to their voiceless counterparts. In order to explore the relationship between voicing and the affective sound of words, we defined a phonetic cue based on the relative proportion of voiced consonants to all consonants in a word. This phonetic cue of voicing was significantly correlated with sound intensity, $r = 0.38$ ($P < 0.0001$), and also with both measures of affective sound: AS-R_{aro}: $r = -0.51$, and AS-R_{val}: $r = 0.49$, as well as AS-P_{aro}: $r = -0.57$, and AS-P_{val}: $r = 0.62$, (all $Ps < 0.0001$).

These results indicate that voiceless consonants sound on average more arousing and negative than voiced consonants, which, in turn, appear to make words sound softer and more pleasing.

6.9.3 Plosive Consonants

Among words with the lowest sound intensity, many include plosive consonants (e.g., Gift /g ɪ f t/ "*gift*", Stich /ʃ t ɪ ç/ *"stab"*). The interruption and explosive release of the air stream in the pronunciation of plosive sounds can be associated with a higher level of arousal, but at the same time, during a stop closure, there is very little acoustic energy. This may explain the lower level of sound intensity (and a higher level of arousal at the same time) for words that include this type of phonemes (see Figure 6.4-A).

Similar to voicing, we defined a phonetic cue indicating the relative proportion of plosive consonants to all consonants in a word. This phonetic cue was significantly correlated with sound intensity: $r = -0.26$ ($P < 0.0001$) and with both affective sound measures, AS-R_{aro}: $r = 0.2$, and AS-R_{val}: $r = -0.16$, as well as AS-P_{aro}: $r = 0.17$, and AS-P_{val}: $r = -0.19$, (all $Ps < 0.0001$), reflecting that while plosives reduce sound energy, they also play a significant role in making the sound (moderately) more negative and arousing.

6.9.4 Hissing Sibilants

In addition to sound intensity and in line with previous findings on vocal expression of emotion (Juslin & Laukka, 2003; Sauter et al., 2010; Scherer, 2003), first formant (F1) and spectral centroid (CoG) appeared to be the dominant features explaining the largest part of variance in words' affective sound, showing a significant effect even at the level of direct ratings for words' affective meaning. A larger high-frequency energy and raising of the first formant are typical characteristics of hissing sibilants (alveolar fricatives and affricates, e.g., /s/, /z/, /ʃ/) which are strongly stressed consonants produced by a high-velocity jet of air against the teeth (see 3B). This results in a literally high-arousing hissing sound, which may account for the cross- and paralinguistic use of these sounds for attracting the attention of others (e.g., "psst!") as well as for their prominent deployment in literature as a stylistic device for cacophony.

Similarly, it is the presence of such a hissing sound following a short vowel that makes the small, but striking difference at the phoneme level between two words referring to one and the same concept from a very basic domain of physical human experience, out of which one is considered rather vulgar and rude, while the other seems more childish and polite: '*piss*' vs. '*pee*'.

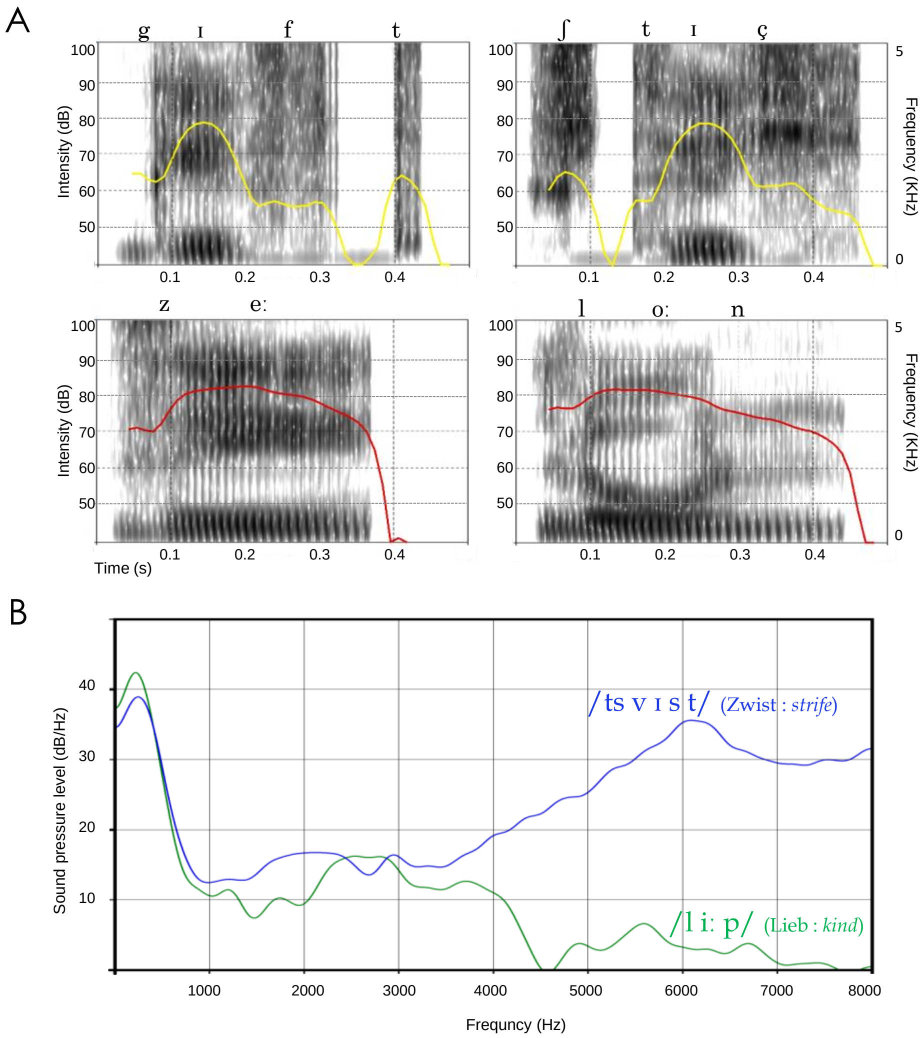


Figure 6.4. A) The time course of sound intensity for the words "Gift /g ɪ f t/ *(gift)*" and "Stich /ʃ t ɪ ç/ *(stab)*" (top, yellow lines) compared to their counterparts "See /z e:/ (*lake*)" and "Lohn /l oː n/ (*wage*)" (bottom, red lines). Short vowels, plosives, and voiceless consonants (as in "Gift" and "Stich") possess smaller integrals of sound energy, whereas sustained high amplitude (see red lines) results in larger sound intensity. This relationship between phonetic features and sound intensity, together with the relationship between sound intensity and 'affective sound' of words, explains the harsh sound of words containing short vowels, plosives, and voiceless consonants. **B)** Spectral analysis shows that hissing sibilants in a word increase the sound's center of gravity (i.e., the magnitude-weighted mean of the frequencies present in the signal), which makes words including this category of phonemes sound harsh and negative (blue line Zwist /ts v ɪ s t/ *(strife)* vs. green line Lieb /l iː p/ *(kind)*).

6.10 General Discussion

The present data demonstrate that words' affective meaning, as reflected in evaluative ratings, is co-determined by words' acoustic-phonetic features. Overall, the results of our computational approach and acoustic analyses, together with the data from the behavioral studies, provide strong support for the hypothesis that phonemes possess affective potential based on their spectro-temporal acoustic features known to modulate emotional vocalization. These results emphasize the iconic nature of the relationship between the (implicit) sound of a phoneme and its affective quality on the one hand and affective meaning of words comprising these phonemes on the other.

As outlined in the introduction, with this study we addressed three major issues generally involved in previous research on iconicity. First, by focusing on a representative number of real words—instead of pseudowords—, our novel results improve the understanding of the effect of implicit sound on the process of meaning making for natural words, in particular concerning their affective meaning. We showed that not only specific sound profiles of words have an affective quality, but also that this quality implicitly influences language users in their final emotional judgment about the meaning of words. Secondly, our behavioral studies and acoustic analyses helped to overcome a major limitation of previous work showing that the relation between affective sound and meaning reflects more than just some statistical regularities within the language (i.e., systematicity) to which language users might be sensitive. Rather, our data suggest that the sound shape of words possesses an inherent affective quality (i.e., iconicity) based on acoustic features that are known to modulate nonverbal emotional communication. Finally, investigating the role of affect and affective meaning of words, we moved beyond the narrow focus on single, limited semantic concepts (see also Monaghan et al., 2014), which enabled us to test sound-meaning correspondences across a wide range of words, representative of the entire lexicon.

Importantly, the iconic affective potential of phonemes (i.e., PAP), as suggested by our results, contributes to the process of affective meaning making even when words are visually presented and silently read. Note that visual word recognition generally involves the activation of phonological codes (Braun et al., 2009; Breen, 2014; Ziegler & Jacobs, 1995) and language users appear implicitly influenced by affective sound of visually presented words when evaluating the affective meaning of these words.

6.10.1 Valence vs. Arousal

Overall, our results were generally stronger for arousal than for valence. This finding aligns with a number of studies on the acoustic properties of emotional speech and hence provides support for an "acoustic arousal dimension". That is acoustic speech properties provide vocal cues to the level of arousal, above that of valence (Aryani et al., 2016; Bachorowski, 1999; Juslin & Laukka, 2003; Sauter et al., 2010). Reviews of earlier attempts to decode emotional significance from vocal cues commonly failed to identify a set of vocal features that reliably differentiate between the levels of valence. Arousal is generally understood as related to a physiological state of being reactive to a stimulus, and it appears plausible that this could be reflected in the vocal behavior of the sender and thus extend to acoustic features of the speech signal. Valence, on the other hand, involves higher order, cognitive, and evaluative processes that are less likely to be detectable at such a basal sublexical level (Aryani et al., 2016; Briesemeister, Kuchinke, & Jacobs, 2014; Briesemeister, Kuchinke, Jacobs, & Braun, 2015; Kuhlmann, Hofmann, Briesemeister, & Jacobs, 2016), thus corresponding less well than arousal to any consistent acoustic mapping.

6.10.2 Phonetic Features

Our analyses of acoustic cues revealed some specific phonetic features as potential candidates of carrying the effect of sound on meaning. Short vowels, compared to their long counterparts, can make words sound more negative and arousing. Also, voiceless consonants, hissing sibilants, and—to some extent—plosives, can significantly contribute to making a word more negative and arousing – as our data suggest, both at the level of sound and perceived meaning.

It is worth noting that these phonetic cues may not be universal across different languages, as not all languages display systematic variations of some of the phonetic features that we focused on (e.g., long/short vowels). Users might rely on different phonological/acoustic affective cues in different languages depending on their phonemic inventory and phonotactic rules - to be investigated in future research on the topic.

6.10.3 Measuring the Affective Sound of Words

Our two studies present, for the first time, two different methods for assessing words' affective sound that can be used in future studies investigating the interaction of words' affective sound and meaning. The poor ICC values for the first method (Study 2a) indicate the difficulty of subjective judgments of the implicit sound of a visually presented word independently of its meaning. Even though we attempted to decontaminate these rating values from the effects of semantic content, this method possesses serious limitations and the poor ICC values call for cautious interpretation of these results.

In contrast to the first method, by using pseudoword material in auditory form in the second study (study 2b), we could largely overcome the limitations of the first approach and provide a better way for assessing words' affective sound, as indicated by the considerably larger ICC values for the pseudoword ratings. Thus our approach based on pseudowords may represent a reliable proxy for words' affective sound in future research.

6.10.4 Alternative Interpretation

The present approach aimed at describing the relation between words' phonology and affective ratings in most basic ways, but our findings might also fit well into proposals concerning iconicity and the organization of the vocabulary: Rather than reflecting a direct, forward influence of acoustic features on affective ratings, PAVs, determining PAPs for words in our data might instead, reflect the systematic occurrence of specific phonemes in words of specific affective meaning (in terms of arousal and valence levels) across the vocabulary of a language. This is because PAVs are is computed as the average of affective ratings of words comprising a given phoneme. In that case, our data establishing close relations between PAP (or PAV) and acoustic features would help explain an apparent systematic distribution of phonemes across the vocabulary as a function of semantic affective values of words: An iconic relation would link affective attributes of the percept or the basic linguistic sign at the phoneme level with affective semantic meaning at the lexical level - adding an internal to the external relation between the signifier and the signified that would have contributed to the evolution of the vocabulary according to affective iconicity.

6.10.5 Limitations and Future Research

Our study is the first to demonstrate an association between affective sound and meaning for real words and across a language lexicon. While providing important novel evidence, it also has limitations future research may attempt to overcome.

When modeling our alternative hypothesis H1, for the sake of simplicity and in the absence of a theoretically or empirically justified theory, we opted for a simple additive method (see Eq. 1). It is, however, possible that words' *Semantic Content* and PAP have an interactive effect on ratings of affective meaning. Similarly, the role of each phoneme in a word for contributing to the PAP might be differentially weighted depending on its position in the word (see Eq. 2). Applying more sophisticated methods such as machine-learning-based regressors (e.g., Jacobs & Kinder, 2018; Jacobs, Schuster, Xue, & Lüdtke, 2017) might help integrate the large number of potentially influential factors into more complete and accurate models of the process of evaluative rating.

Also, a number of the acoustic features we used are measured as average frequencies, which precludes the use of dynamic sound features (e.g., spectral flex). Employing other methods based on dynamic changes of the sound signal might increase the accuracy of acoustic models predicting ratings of words' affective sound. A more sophisticated approach, for instance, might use the matrix of the spectrogram to quantitatively represent the sound envelope. Since the length of the audio signal (i.e., the length of words or pseudowords) differs for each item, the challenge of such an approach would be to find an appropriate method to classify the (pseudo)words' affective sound based on a series of independent variables, the number of which depends on physical signal length.

Alternatively, our acoustic analysis can be complemented by the use of phonetic categories (e.g., voiced/voiceless, obstruent/sonorant, etc.) to relate these categories to the effect of sound on words' affective meaning. In a simple phonetic approach, each phoneme in a word will represent a vector of phonetic features. Consequently, an entire word—comprising different phonemes—can be described as a concatenation of vectors of phonetic features, which can be used to calculate the contribution of any phonetic feature to the affective sound of words; in a similar fashion to our approach concerning PAVs. A practical approach concerning the use of phonetic features instead of acoustic variables would be the use of phonological cues defined as the proportion of consonants with particular manner and place features, and the average height and position of vowels (as provided in Monaghan, Christiansen, & Chater, 2007). These cues can be used in the same way as our acoustic variables to identify phonological features underlying the PAPs. An advantage of this method would be the simple classification of the phonological construction of a word and its contribution to the sound to meaning relation. Our initial investigation has shown that such phonological cues can account for a significant portion of variance in the PAPs (25% for arousal, and 15% for valence), with the proportion of voiced consonants, and the average of vowel roundedness in a word being the most important predictors for both models of arousal and valence.

Another approach for measuring the affective sound of words can make use of the insights of sonority theory (Clements, 1990; Stenneken, Bastiaanse, Huber, & Jacobs, 2005). For this, each word can be assigned a sonority score which may also systematically contribute to affective (and aesthetic) ratings, as supported by recent findings concerning ratings of the aptness of metaphors and the beauty of words (Jacobs & Kinder, 2018; Jacobs, 2017).

6.10.6 Practical Applications

Our findings on the effects of implicit sound on affective meaning, and specifically our acoustic model for measuring the affective sound of words effectively suggest a method for constructing words and pseudowords associated with specific affects (positive/negative, arousing/calming) or emotions (e.g., fear, disgust), which can have broad applications in various contexts from marketing and advertising to art and literature. For instance, in the field of product and brand naming, previous work has shown that the sound of a product's name can in general set and modify consumer expectations about the likely attributes of the products (Schrott & Jacobs, 2011; Spence, 2012) and that names with negative sounds were least preferred regardless of product category (Baxter & Lowrey, 2011). Here, our method for assessing the affective sound of words based on its acoustic features could provide a substantial improvement to previous work, which was usually based on the manipulation of a limited group of sounds (e.g., front vs. back vowels). Likewise, in artistic contexts, such as film, literature, and in particular, poetry, our method could be applied to evoke and verify particular emotional effects by use of words that possess specific implicit affective sounds.

Poetry is probably the best example of sound meaning interaction: while it is inherently concerned with the expression and elicitation of emotions (Jacobs, 2015a; Lüdtke, Meyer-Sickendieck, & Jacobs, 2014; Schrott & Jacobs, 2011), it is deeply rooted at the perceptual level in the domain of sound (Jacobs, Lüdtke, et al., 2016; Kraxenberger & Menninghaus, 2016; Schrott & Jacobs, 2011; Ullrich et al., 2017). Indeed, poetry has always artfully deployed sound patterns to shape order, to create a new layer of meaning, and to emphasize the affective meaning in a text. With the present study we provide a complementary method to previous attempts for analyzing poetic texts at the sublexical level (Aryani et al., 2013, 2016; Auracher, Albers, Zhai, Gareeva, & Stavniychuk, 2011; Kraxenberger & Menninghaus, 2016; Ullrich et al., 2017; Whissell, 1999, 2000), and for further examination of the influence of sound structure on affective and aesthetic reactions to verbal material intended to elicit a certain emotional impact in readers, such as advertisements, political speeches or manifests.

6.11 Conclusion

The present studies provide novel results on the contribution of the implicit sound of a word to its affective meaning. Our findings have the potential to shed new light on various unanswered questions regarding the evolution, organization, and processing of human language by drawing attention to the role of affect as well as by substantiating the psychological reality of iconicity in everyday language. These new insights may pave the way for further cross-linguistic investigations, as well as the detailed study of the neural substrates underlying the effect of phonology and sound-meaning interaction in language use; a phenomenon creatively exploited particularly by Poe and other poets throughout history.

Chapter 7

The Sound of Words evokes Affective Brain Responses[3]

7.1 Abstract

The long history of poetry and the arts, as well as recent empirical results suggest that the way a word sounds (e.g., soft vs. harsh) can convey affective information related to emotional responses (e.g., pleasantness vs. harshness). However, the neural correlates of the *affective potential* of the sound of words remain unknown. In an fMRI study involving passive listening, we focused on the affective dimension of arousal and presented words organized in two discrete groups of sublexical (i.e., sound) arousal (*high* vs. *low*), while controlling for lexical (i.e., semantic) arousal. Words sounding high arousing compared to their low arousing counterparts, resulted in an enhanced BOLD signal in bilateral posterior insula, the right auditory and premotor cortex, and the right supramarginal gyrus. This finding provides first evidence on the neural correlates of affectivity in the sound of words. Given the similarity of this neural network to that of nonverbal emotional expressions and affective prosody, our results support a unifying view that suggests a core neural network underlying any type of affective sound processing.

[3] This chapter is published as: Aryani, A., Hsu, C. T., & Jacobs, A. M. (2018). The Sound of Words Evokes Affective Brain Responses. *Brain Sciences*, *8*(6), 94.

7.2 Introduction

When communicating, humans usually express emotion through two different signaling systems: verbal vocalization, i.e., relating the semantic content of particular phoneme combinations (words), and nonverbal vocalization, i.e., relating paralinguistic cues such as intonation or rhythm. According to this perspective of division, there is no inherent relevant information in phonemes per se (Saussure, 1916/2011). Rather, affective information in speech is conveyed either through conventional and arbitrary sound-meaning mappings or through the prosodic features of a vocalization.

However, the long history of poetry, as the most ancient record of human literature, as well as recent empirical results suggest a possible connection between phonemes and another layer of affective meaning beyond the conventional links (Aryani et al., 2013, 2016; Jacobs, 2015a; Jakobson, 1965; Schrott & Jacobs, 2011). Stylistic devices such as euphony or cacophony are instructive examples indicating how the sound of a word can evoke a feeling of pleasantness or harshness, respectively. Children already possess the ability to easily evaluate whether a word sounds positive/negative or beautiful/ugly (Jacobs et al., 2015). This idea has been supported by recent experimental evidence highlighting the role of sound in affective meaning making (Aryani, Conrad, Schmidtcke, & Jacobs, 2018), as well as its contribution to the beauty of words (Jacobs, 2017)

Although the brain networks involved in emotion processing for both verbal and nonverbal stimuli have been well studied, little is known about the neural correlates of the *affective potential* of a word's sound (but see Ullrich et al., 2016, for an ERP study). In the present study, we examined the neuropsychological reality of sublexical sound effect, and aimed at identifying its underlying brain network. To quantify the affectivity of the sound of words we used a recent psycho-acoustic model (Aryani, Conrad, et al., 2018) which is based on a two-dimensional space of valence (ranging from pleasant to unpleasant) and arousal (ranging from calm to excited) (Osgood, 1952; Wundt, 1908). The model relies on the fact that acoustic features characterizing phonemes and their combinations (as in words) are similar to those modulating emotional vocalization and affective prosody (e.g., sound formants, sound intensity). Thus, these specific features extracted from the sound profile of a word can predict *affective potential* of the sound of that word (Aryani, Conrad, et al., 2018). Also, previous studies showed a high similarity of acoustic cues to affective judgments across different types of affective sounds (e.g., speech, music, and environmental sound) (Weninger, Eyben, Schuller, Mortillaro, & Scherer, 2013). Due to this similarity, we hypothesize that affectivity in the sound of a word will be processed in similar brain regions that are involved in processing other types of affective sounds, as proposed by a *unifying neural network* perspective of affective sound processing (Frühholz et al., 2016).

In an fMRI study involving a passive listening task, we presented participants with words varying in their sublexical affectivity (sound) while controlling for lexical (semantic) affectivity. Specifically, we focused on the affective dimension of arousal, as previous studies showed that arousal, compared to valence, can be more reliably decoded and identified from vocal cues (Aryani, Conrad, et al., 2018; Aryani et al., 2016; Juslin & Laukka, 2003; Weninger et al., 2013).

7.3 Materials and Methods

7.3.1 Stimuli

120 nouns (one to three syllables long) were selected for a 2x2 design (30 words for each condition) characterized by an orthogonal twofold manipulation of *lexical* and *sublexical arousal*. For *lexical arousal* we used ratings of words' affective meaning (min = 1: very low arousing, max = 5 very high arousing) from the normative database BAWL-R (Võ et al., 2009). *Sublexical arousal* was calculated based on features extracted from the acoustic representation of words applying the acoustic model developed in our previous work (see study 2b in Aryani, Conrad, et al., 2018). For this, words were uttered in a list-like manner by a professional male actor who was a native speaker of German and recorded with a sampling frequency of 48 kHz

and 16 bits per sample. Audio files were then normalized to have the same loudness by matching their root-mean-square (RMS) power. Words were divided into two distinctive conditions of “high” and “low” arousing for each of the factors *lexical arousal* (‘High’ > 3.25, ‘Low’ <2.75) and *sublexical arousal* (‘High’ > 3, ‘Low’ < 3), and carefully controlled for relevant psycholinguistic variables across all of four cells of experimental conditions. *Lexical arousal* (and *lexical valence*) was closely controlled for between the two cells of *sublexical arousal*, and vice versa (Table 7.1). In order to create an acoustic baseline, we randomly selected 16 words from the word material (4 from each condition) and converted them to signal-correlated noise (SCN). Along with our stimulus material (120 words + 16 SCN), a total of 76 additional words (mostly emotionally neutral) were presented which were a part of another study, and were discarded from further analysis here.

Variable	Word Category								Inferential Statistics
	HH		HL		LH		LL		
	M	SD	M	SD	M	SD	M	SD	
Lexical Arousal	4.07	0.24	4.04	0.22	1.99	0.16	1.99	0.18	$F(3,116)=983$, $p<0.0001$
Lexical Valence	-1.83	0.52	-1.83	0.51	0.22	0.36	0.18	0.37	$F(3,116)=205$, $p<0.0001$
Sublexical Arousal	3.36	0.31	2.76	0.19	3.30	0.27	2.77	0.21	$F(3,116)=50.5$, $p<0.0001$
Word Frequency	0.64	0.75	0.74	0.76	0.57	0.78	0.51	0.75	$F(3,116)=0.47$, $p=0.69$
Imageability	4.78	1.01	4.56	1.0	4.93	0.90	5.02	1.16	$F(3,116)=1.11$, $p=0.34$
Syllables	1.86	0.73	2.1	0.54	2.0	0.69	2.03	0.61	$F(3,116)=0.68$, $p=0.56$
Phonemes	5.3	1.36	5.23	1.10	5.13	0.89	4.93	1.20	$F(3,116)=0.57$, $p=0.63$
duration (ms)	873	116	850	102	826	108	836	100	$F(3,116)=1.06$, $p=0.36$

Table 7.1. Characteristics of word stimuli. HH= High-High, HL=High-Low, LH=Low-High, LL=Low-Low: the first letter indicates the *lexical* and the second *sublexical arousal*.

7.3.2 Participants

Twenty-nine right-handed German native speakers (17 women, mean age 25.2 years, range: 20-35 years) with no history of neurological or psychiatric illness or any hearing problems volunteered to participate in the study, receiving either 15 Euros or psychology course credit for their participation. Handedness was determined using the Edinburgh Inventory (Oldfield, 1971).The Ethical Committee of the Freie Universität Berlin had approved the investigation. Informed consent was obtained according to the Declaration of Helsinki.

7.3.3 Procedure

Spoken words were presented via MRI-compatible headphones sufficiently shielded from scanner noise to ensure clear perceptibility. Participants were instructed to pay attention and to carefully listen to the words. A trial began with the presentation of a fixation cross for between 1500 ms and 6500 ms, jittered in steps of 500 ms, in the center of the screen. Jittering durations and the stimulus presentation order over different experimental conditions (HH, HL, LH, LL, SCN, Fillers), were optimized to ensure a maximal signal-to-noise ratio. After presentation of a stimulus the fixation cross disappeared. All blocks were set to a fixed length of 370 volumes. A total number of 10 trial words were presented prior to the experiment, which were excluded from the analysis. Words were split and presented in two runs. Between the two runs the participants could take a break.

7.3.4 fMRI data acquisition

Imaging data were collected on a Siemens Tim Trio 3T MR scanner. Functional data used a T^{*}_{2}-weighted echo-planar sequence [slice thickness: 3 mm, no gap, 37 slices, repetition time (TR): 2 s, echo time (TE): 30 ms, flip angle: 70°, matrix: 64 × 64, field of view (FOV): 192 mm, voxel size: 3.0 mm × 3.0 mm × 3.0 mm, 2 x 305 volumes, acquisition time: 2x 610 s]. At the beginning of the experimental session, magnitude and phase images for the field map were acquired: [slice thickness: 3 mm, no gap, 37 slices, TR: 488 ms, 2 TE: 4.92 and 7.38 ms, flip angle: 60°, matrix: 64 × 64, FOV: 192 mm, voxel size: 3.0 mm × 3.0 mm × 3.0 mm, acquisition time: 65 s]. Individual high-resolution T1-weighted anatomical data (MPRAGE sequence) were also acquired (TR: 1.9, TE: 2.52, FOV: 256, matrix: 256 × 256, sagittal plane, slice thickness: 1 mm, 176 slices, resolution: 1.0 mm × 1.0 mm × 1.0 mm).

7.3.5 Post-scan tests

Unannounced recognition test: At the end of the experiment, outside the scanner, an unannounced recognition test was performed to assess participants' involvement in the task and mnemonic effects of the experiment. Participants were presented with the same 120 words used in the scanner (OLD) mixed with 120 new words (NEW) which were matched with OLD items for word frequency, number of letters, number of phonemes, number of syllables, and imageability rating, as well as valence and arousal (selected from the same range as used for OLD items). Participants were asked to rate how confident they were that the presented word was or was not part of the word list in the scanner (from certainly not presented in the scanner = 1 to certainly presented in the scanner = 5).

Ratings: After the recognition test, in two separate rating studies, participants were asked to evaluate the words presented in the scanner for their *lexical arousal* (study1) and *sublexical arousal* (study2). For the latter, participants were instructed to only concentrate on the sound aspect of the words while trying to suppress their meaning (cf. Aryani, Conrad, et al., 2018).

7.3.6 fMRI Preprocessing

The fMRI data were preprocessed and analyzed using the software package SPM12 (www.fil.ion.ucl.ac.uk/spm). Preprocessing consisted of slice-timing correction, realignment for motion correction, magnetic field inhomogeneity correction through the creation of a field map, and coregistration of the structural image onto the mean functional image. The structural image was segmented into gray matter, white matter, cerebrospinal fluid, bone, soft tissue, and air/background (Ashburner & Friston, 2005). A group anatomical template was created with DARTEL (Diffeomorphic Anatomical Registration using Exponentiated Lie algebra; Ashburner, 2007) toolbox from the segmented gray and white matter images. Transformation parameters for structural images were then applied to functional images to normalize them to the brain template of the Montreal Neurological Institute (MNI) supplied with SPM. Functional images were resampled to a resolution of 1.5 × 1.5 × 1.5 mm, and spatially smoothed with a kernel of 6 mm full-width-at-half-maximum during normalization.

7.3.7 fMRI Analysis

Voxel-wise fixed effects contrast images made by subtraction analyses were performed at the single subject level and random effects analyses (Holmes & Friston, 1998) were conducted at the group level to create SPM contrast maps. On the single-subject level, each of the six conditions (HH, HL, LH, LL, SCN, and FILLERS) was convolved with the haemodynamic response function (HRF). Events were modeled as delta functions with zero duration. The beta images of each conditional regressor were then taken to the group level, where a full-factorial 2nd level analysis with the factors *lexical arousal* and *sublexical arousal* was used. An unconstrained non-directional 2x2 ANOVA whole brain analysis was performed with the factors *lexical arousal* (High, Low) and *sublexical arousal* (High, Low), to investigate the overall presence of main and interaction effects. For whole-brain fMRI analyses, we used the cluster defining threshold (CDT) of $p < 0.005$, then applied cluster-level family-wise error (FWE) correction to $p < 0.05$ for the entire image volume, as suggested by Liebermann and Cunningham (Lieberman & Cunningham, 2009) for studies in cognitive, social and affective neuroscience. The labels reported were taken from the 'aal' labels in the WFU Pickatlas Tool. The Brodmann areas (BA) were further checked with the Talairach Client using nearest gray matter search after coordinate transformation with the WFU Pickatlas Tool.

7.4 Results

7.4.1 Behavioral Results

<u>Recognition Test:</u> Across all participants, we performed a Linear Mixed Model analysis predicting the recognition rate, with word category (OLD vs. NEW) as fixed factor and words as well as participants as random factors. Results supported a performance above chance for recognizing OLD words, with a significantly higher score average ($M = 3.53$) compared to NEW words ($M = 2.54$): $t = -20.6$, $p < .0001$. We next

performed simple t-tests to compare the recognition rate between the levels of word category (OLD vs. NEW) separately for each participant. An effect of word category (OLD vs. NEW) on accuracy was observed for 27 participants out of 29 ($t = 6.4 \pm 3.2$). These results indicate that the majority of participants had been attentive during the passive listening task. Two participants with a performance not higher than chance level ($t = 0.28$, $t = 1.14$) were consequently excluded from further analyses.

<u>Ratings:</u> To check the reliability of our experimental manipulations, we correlated the rating values for *lexical* and *sublexical arousal* used for the experiment with our post-scan data. For both, the coefficients were very high: $r = 0.97$, $p < .0001$, (r_{min} among all participants = 0.73), and $r = 0.76$, $p < .0001$ (r_{min} among all participants = 0.49), respectively (Figure 7.1).

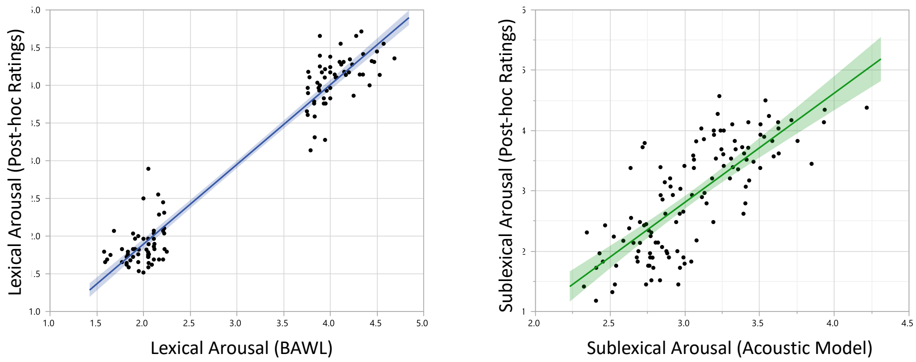


Figure 7.1. Results of post-scan ratings were highly correlated with affective measures used for the fMRI-experiment. Left: lexical arousal ($r = .97$), Right: sublexical arousal ($r = .76$).

7.4.2 Neuroimaging Results

<u>Main effect of all words compared to SCN:</u>

The comparison between all words contrasted with the baseline condition of the SCN revealed left-lateralized activations in core language areas, i.e., the inferior frontal gyrus (IFG), middle and superior temporal gyrus, and inferior parietal lobule

(BA 40), suggesting that this experiment successfully tapped into the language processing system. Activity was also observed in bilateral parahippocampal gyrus, middle frontal gyrus and precentral gyrus, as well as the left superior frontal gyrus, the fusiform area, the right caudate, and superior parietal lobule.

Main effect of the category lexical arousal:

Words with higher levels of lexical arousal (Lex H > Lex L) elicited a large cluster of activation in the left and right dorsolateral and medial prefrontal cortex, a cluster of activation extending from the left IFG into the anterior end of left temporal lobe, as well as a cluster including the left posterior cingulate cortex (PCC) and precuneus (Table 7.2, Figure 7.2). Words with lower level of lexical arousal (Lex L > Lex H) elicited a cluster of activation in the left extrastriate cortex in middle occipital gyrus (BA 19) extending to the fusiform area (BA 37) and mirrored by a smaller cluster in the right occipital lobe (BA 37), as well as a cluster of activation immediately posterior to the primary somatosensory cortex (BA 5).

Contrast			MNI coordinates			Z	K
	Anatomical Definition		x	y	z		
LexH>LexL	L/R	Medial Frontal Gyrus (BA 9)	-3	56	20	5.12	4079
	L	IFG (BA 47), Temporal Pole (BA 38)	-30	21	-17	4.48	672
	L/R	Cuneus, Precuneus (BA 7, BA31)	-3	-68	32	4.01	694
	L	Posterior Cingulate Cortex (BA 23)	-8	-47	26	3.90	492
LexL>LexH	L	Middle Occipital Gyrus (BA 37, 19)	-53	-60	-11	5.88	1244
	R	Middle Occipital Gyrus (BA 37)	56	-57	-8	3.88	515
	L	Somatosensory Cortex (BA 5)	-21	-47	54	4.39	717
SubH>SubL	L	Posterior Insula (BA 13)	-42	-15	-1.5	4.86	861
	R	Posterior Insula (BA 13)	39	-15	1.5	4.78	943
	R	Superior Temporal Area (BA 40, BA 22)	51	-38	24	4.58	852
	R	Supplementary and Premotor area (BA 6)	12	-6	54	3.87	524

Significant peak voxel for all comparisons at $p < 0.05$ FWE-corrected: H = High arousal, L = Low Arousal, Lex = Affective lexical meaning of arousal, Sub = Affective sublexical sound of arousal. MNI = Montreal Neurological Institute, L/R = Left/Right

Table 7.2. Results for two main contrasts of *lexical* and *sublexical* arousal.

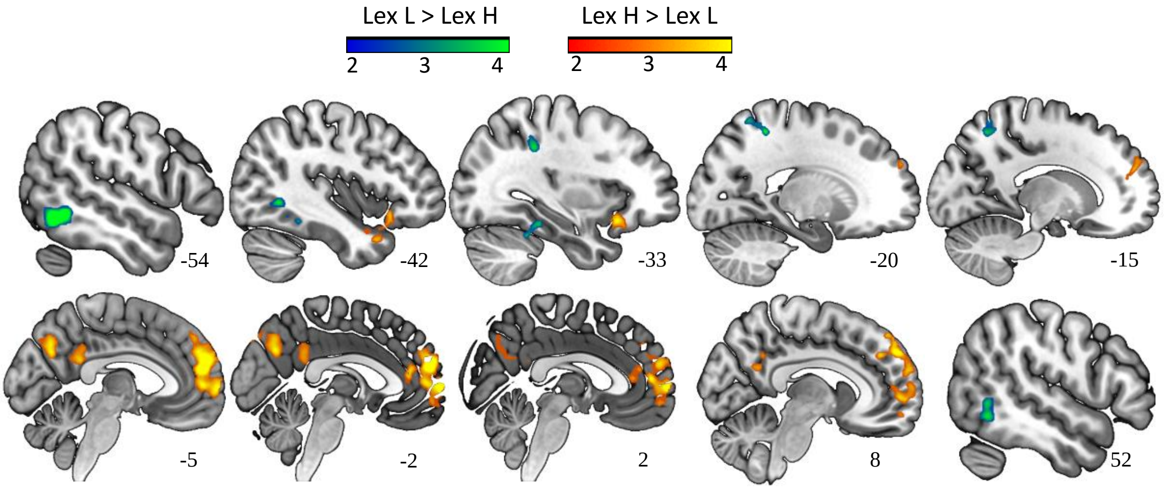


Figure 7.2. Words with a higher degree of lexical arousal (Lex H > Lex L) elicited stronger activation in a widespread network of medial and inferior frontal gyrus, as well as temporal pole, cuneus, precuneus, and posterior cingulate cortex. The reverse contrast (Lex L > Lex H) resulted in an enhanced BOLD signal in visual and somatosensory cortex ($p < 0.05$, FWE-corr.).

<u>Main effect of the category sublexical arousal:</u> Words with higher *sublexical arousal* (Sub H > Sub L) evoked an increased BOLD signal in bilateral posterior insula, a cluster including the posterior part of superior temporal area and the right supramarginal gyrus, as well as the right premotor cortex and supplementary motor area (Figure 7.3). No activation was observed for the contrast Sub L > Sub H.

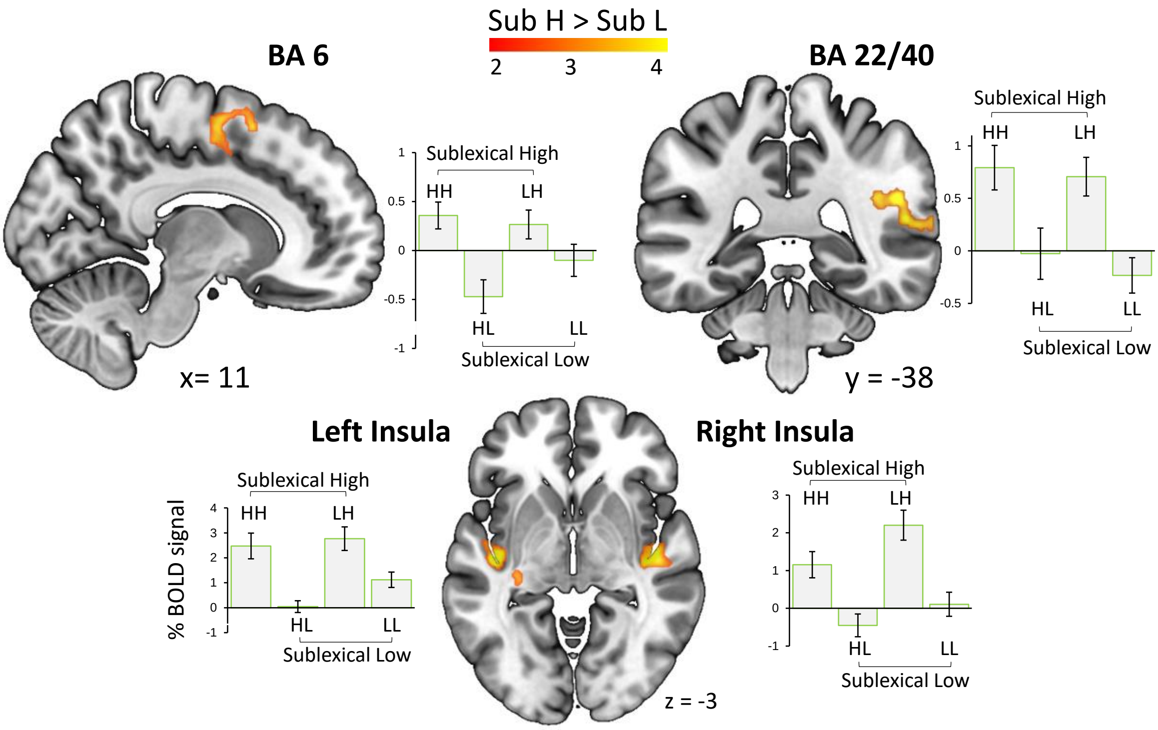


Figure 7.3. The main effect of *sublexical arousal* (i.e., words sounding high vs. low arousing) and the related pairwise comparisons were associated with an enhanced BOLD signal in bilateral posterior insula, superior temporal cortex (BA 22 extending to BA40), as well as supplementary and primary motor cortex (BA 6) ($p < 0.05$, FWE-corr.).

7.5 Discussion

The current study investigated the neural correlates underlying the *affective potential* of a word's sound and whether brain regions involved in processing emotional vocalization and affective prosody are also used to process affectivity in the sound of a word.

The overall activation observed for the effect of *lexical arousal* (Lex H > Lex L) is in accordance with previous findings showing the involvement of dorsolateral and medial prefrontal cortex, as well as PCC, LIFG, and temporal pole in appraisal and general processing of affective stimuli (Etkin, Egner, & Kalisch, 2011; Kuchinke et al., 2005; Kuhlmann et al., 2016; Lewis, Critchley, Rotshtein, & Dolan, 2007). On the other hand, in the inverse contrast, i.e., Lex L > Lex H, activations of visual and somatosensory areas were observed, suggesting a stronger involvement of perceptual- and image-based systems for processing less emotional words. That is, the semantic processing of words with a lesser emotional connotation is embodied mostly in the brain systems devoted to sensory information about physical word experiences, whereas emotion words are more anchored in affective experiences. This finding is in line with the theories of embodied language stating that concepts are formed as a result of interactions with the real world in various sensory, motor and affective information about external world experiences (e.g., Gallese & Lakoff, 2005; Glenberg, 2010; Jacobs, Hofmann, et al., 2016; Vigliocco et al., 2009).

By replicating the results of previous studies for both contrasts, Words >SCN (see Results) and Lex H > Lex L, as well performing an unannounced recognition test, we showed that the present experiment successfully engaged participants in carefully listening to words, thus assuring the reliability of the results including those of the subsequent effect of *sublexical arousal*. Results for the main effect of *sublexical arousal* (Sub H > Sub L) indicate a substantial sharing between the processing networks for the affectivity in the sound of words and other types of affective sounds. This provides the first neuroimaging evidence for the emotion potential lying in the sound of words, and, importantly, it supports the idea of a *unifying neural network* of affective sound processing rather than a traditional view that proposes distinct neural systems for specific affective sound types (Frühholz et al., 2016). According to this view, all affective sounds consistently induce brain activity in a common core network which consists of i) superior temporal cortex and amygdala: likely involved in decoding of affective meaning from sound with amygdala's involvement rather in less complex stimuli, ii) frontal and insular regions: likely involved in the evaluation and perception of sound, respectively, and iii) motor-related areas: likely involved in emotional behavior (Frühholz et al., 2016).

The observed activation in the right superior temporal area (BA 22) has been associated, for instance, with intensity of both happy and angry intonations (Ethofer et al., 2006). This effect may be driven by a combination of acoustic features expressing the arousal in the speaker's voice (Wiethoff et al., 2008). Superior temporal areas have been shown to be involved in discriminating sound pitch and sound intensity (Belin et al., 1998) which are two acoustic features shaping affective prosody (Juslin & Laukka, 2003; Dirk Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006). Crucially, these two features serve as significant predictors in the acoustic model of *sublexical arousal* (Aryani, Conrad, et al., 2018) used in the present study. The absence of the activation of amygdala in this part of network may indicate the complexity of speech signals, and is in line with previous findings that show amygdala's involvement in the processing of less complex affective sounds (e.g., non-human environmental sounds, and nonverbal vocalizations), probably due to their function as an emotional signal at a very basic level (Frühholz, Trost, & Grandjean, 2014; Frühholz et al., 2016a).From the expected response in fronto-insular brain system, we observed significant clusters of activation in bilateral insula, but no activation in any of the frontal regions. Concerning the widespread connections of the posterior insula with the auditory cortex and many afferents that it receives from thalamus, previous reports have shown the insula's significant involvement in auditory temporal processing of most types of emotional sound (Frühholz et al., 2016; Mirz, Gjedde, Sødkilde-Jrgensen, & Pedersen, 2000; Trost, Ethofer, Zentner, & Vuilleumier, 2012). Insula has also been proposed to function as a mediator between sensory and affective brain systems in the perception of affective sounds, thereby enabling a self-experience of emotions in terms of a subjective feeling (Frühholz et al., 2016a; Mirz et al., 2000). In regard to the anticipated response in frontal brain regions (e.g., IFG), the absence of such an activation in our study is presumably due to the lack of affective evaluations in the experimental task we used: that is, passive listening. Increasing activation in IFG, as well as its connectivity with STG, is associated with evaluative judgments of affective prosody (Leitman, 2010) which our participants were not asked for (but see (Frühholz & Grandjean, 2012) for a refined fronto-temporal network for the decoding of affective prosody).

In line with the proposed view of a unifying core network, we also observed a cluster of activation in premotor cortex and supplementary motor area. This finding aligns with reports on motor responses to the variety of high arousing sounds (Löfberg, Julkunen, Pääkkönen, & Karhu, 2014; Zald & Pardo, 2002) suggesting that emotionally charged stimuli mobilize the motor system to be prepared to take action for approach or withdrawal. This sound-motion relationship has also been proposed to underlie the feeling of being in the 'groove' (Janata, Tomic, & Haberman, 2012), or a general urge to move when listening to music (Trost et al., 2012).

7.6 Conclusion

Our study is the first attempt to understand the brain response to the *affective potential* lying in the sound of words. In accordance with a *unifying neural network* view for affective sound processing, we observed BOLD responses in superior temporal area, insula, and premotor cortex, suggesting that the affectivity in the sound of words shares a processing network with other types of emotional vocal cues. Our study thus provides first neuroimaging evidence for a phenomenon that has long been deployed in poetry and the arts, i.e., evoking affective (and aesthetic) responses by the use of certain words with specific sound patterns. Our data also suggests that human subjects are sensitive to the affective information in the sound of words even when the attentional focus is not directed on that aspect.

Chapter 8

Affective congruence between sound and meaning of words facilitates semantic decision[4]

8.1 Abstract

A similarity between the form and meaning of a word (i.e., iconicity) may help language users to more readily access its meaning through direct form-meaning mapping. Previous work has supported this view by providing empirical evidence for this facilitatory effect in sign language, as well as for onomatopoetic words (e.g., cuckoo) and ideophones (e.g., zigzag). Thus, it remains largely unknown whether the beneficial role of iconicity in making semantic decisions can be considered a general feature in spoken language applying also to "ordinary" words in the lexicon. By capitalizing on the affective domain, and in particular arousal, we organized words in two distinctive groups of iconic vs. non-iconic based on the congruence vs. incongruence of their lexical (meaning) and sublexical (sound) arousal. In a two-alternative forced choice task, we asked participants to evaluate the arousal of printed words that were lexically either high or low arousing. In line with our hypothesis, iconic words were evaluated more quickly and more accurately than their non-iconic counterparts. These results indicate a processing advantage for iconic words, suggesting that language users are sensitive to sound-meaning mappings even when words are presented visually and read silently.

[4] This chapter is published as: Aryani, A., & Jacobs, A. M. (2018). Affective Congruence between Sound and Meaning of Words Facilitates Semantic Decision. *Behavioral Sciences*, 8(6), 56.

8.2 Introduction

Classic linguistic approaches to meaning embed a core assumption that the way a word sounds does not play any contributing role in its meaning (Saussure, 1916). Rather, language users would access the meaning of words solely through learned, and per se, arbitrary links between linguistic symbols and their cognitive representations. Recent findings, however, support a more differentiated view by acknowledging the importance of non-arbitrary sound-meaning mappings in language processing and in the organization of vocabulary (see Dingemanse et al., 2015; Perniss & Vigliocco, 2014; Schmidtke, Conrad, et al., 2014, for reviews). These findings distinguish between two types of motivations for such sound-meaning mappings (Dingemanse et al., 2015): iconicity, which is based on similarities between aspects of sound and aspects of meaning (e.g., onomatopoeia), versus systematicity, which is based on statistical regularities in language that link specific patterns of sound to specific semantic or grammatical concepts (Christiansen & Monaghan, 2016; Farmer et al., 2006; Reilly et al., 2012). However, in many cases, the nature of the relationship between sound and meaning is not particularly clear. The phonaestheme /sn-/ appearing as an initial sound cluster in many English words related to “mouth” or “nose” may serve to illustrate this issue. It is an ongoing debate whether in this case a specific (nasal) quality of the sound of /sn-/links this sound to the concepts of “mouth” or “nose”, or if rather the organization of the vocabulary has evolved in a way so that this specific sound cluster over-proportionally appears in words that are related to these concepts.

In the present study, we aimed at investigating iconicity and its potential facilitatory role in lexico-semantic processing. In addition to a direct acoustic mapping, as in the case of onomatopoeia, iconic words can also evoke other sensory (including visual and tactile), motor, or affective experiences by systematically relating properties of such experiences to phonetic features or acoustic properties (Meteyard, Stoppard, Snudden, Cappa, & Vigliocco, 2015b; Perniss & Vigliocco, 2014) as evident in ideophones (e.g., “twinkle”, Dingemanse et al., 2016; Kwon & Round, 2014; Lockwood et al., 2016), in mimetic words (Childs, 2015; Kita, 1997), or in affective responses associated with the phonology of swear words (Aryani, Conrad, et al., 2018; Bowers & Pleydell-Pearce, 2011). Owing to such a sound-meaning mapping, iconic words have been suggested to be capable of directly evoking sensory, motor, or affective experiences by systematically relating properties of such experiences to phonetic features or acoustic properties of words (Aryani, Conrad, et al., 2018; Meteyard et al., 2015; Perniss & Vigliocco, 2014; Vinson, Thompson, Skinner, & Vigliocco, 2015).

From a learning perspective, empirical evidence for both children and adults support an iconic advantage for learning the vocabulary of a language with which they had no prior experience. For instance, the meaning of Japanese iconic verbs, compared to non-iconic verbs, have been shown to be better learned and generalized by English speaking children (de Ruiter, Theakston, Brandt, & Lieven, 2018; Imai et al., 2008; Kantartzis et al., 2011). These results are in line with the analyses of longitudinal diary data which suggest that over the course of language development iconic words are in general acquired earlier and potentially employed by infants as a bootstrapping mechanism on both lexical and phonological levels (Laing, 2015; Monaghan et al., 2014; Thompson et al., 2012).

By the same token, as in vocabulary learning, iconicity has shown to facilitate language processing. Particularly, in sign languages, in which iconic relationships between form and meaning are far more prevalent than in spoken languages (Elliott & Jacobs, 2014; Taub, 2001), iconicity has been shown to facilitate a variety of language processing tasks such as picture–sign matching, phonological decision, and picture naming (Vinson et al., 2015b), indicating that during lexical processing, iconic words benefit from an additional path between form and meaning by activating conceptual features related to perception and action (see also Thompson et al., 2012). Also, onomatopoetic words imitating animal sounds (e.g., “cuckoo”) have been shown to recruit brain regions involved in the processing of both verbal and nonverbal sounds (Hashimoto et al., 2006). These findings indicate that iconic words profit from additional processing networks that can facilitate both vocabulary learning and lexical processing (Dingemanse et al., 2015; Imai et al., 2008).

Nevertheless, unlike pioneering works on the facilitatory effect of iconicity in sign language (Thompson et al., 2012; Vinson et al., 2015) which also laid the groundwork for the theoretical framework of such investigation, related research on spoken language still faces some limitations. Previous work on the processing advantage of iconicity in lexico-semantic processing of spoken language has so far mainly focused on either nonwords (Parise & Pavani, 2011; Westbury, 2005), onomatopoeia, and ideophones, including Japanese mimetic words (Dingemanse et al., 2016; Iwasaki et al., 2007; Kwon & Round, 2014; Lockwood et al., 2016), or on cases typically considered as systemticity (Bergen, 2004; Farmer et al., 2006; Reilly et al., 2012). Therefore, empirical evidence on whether iconic mappings in a real word can in general facilitate lexico-semantic processing is missing. This is chiefly due to a lack of appropriate measures for both the sound and meaning aspects of words. This limitation has prevented previous research on real spoken words to move beyond onomatopoeia and ideophones, leaving open the question of whether iconicity could be considered a "general" mechanism facilitating language processing. In addition, due to the limited number and the specific properties of onomatopoetic words and ideophones (e.g., phonological construct, frequency, etc.), no empirical research has so far investigated the effect of iconicity on lexico-semantic processing in a carefully controlled experimental paradigm. In the present investigation, we aimed at extending the results of previous works to the facilitatory effect of iconicity in "ordinary" words during a semantic decision task.

By capitalizing on the affective domain, in a recent study, Aryani et al. (Aryani, Conrad, et al., 2018) provided quantitative measures for *lexical* affective meaning and *sublexical* affective sound of words in a two dimensional space of valence (ranging from pleasant to unpleasant) and arousal (ranging from calm to excited), with both measures empirically validated at behavioral and neurobiological levels of analysis (see Jacobs et al., 2015, for the lexical and Aryani, Conrad, et al., 2018; Aryani, Hsu, & Jacobs, 2018, for the sublexical measure). The results of the large-scale lexicon analysis suggest that affectivity in the implicit sound of printed words can influence the listener in their judgment about the words' affective meaning. In the present study, we aimed at extending the scope of the above mentioned work and categorized word in two groups of iconic vs. non-iconic based on the congruence between sound and meaning in the affective domain. We asked whether iconicity can facilitate evaluative decisions on words' affective content: Imagine two words representing similar lexical affective content (e.g., both high arousing), but one sounds harsh (congruent with the meaning) while the other sounds soft and calming (incongruent with the meaning). Which one will be classified more quickly and more accurately as high arousing in a decision task on affective meaning? A null hypothesis (H0), according to the established notion of linguistic arbitrariness (Saussure, 1916), will expect no significant differences, while our alternative hypothesis (H1) predicts iconic (i.e., congruent) words to be evaluated more quickly and more accurately than non-iconic (i.e., incongruent) words (Meteyard et al., 2015; Perniss & Vigliocco, 2014; Vinson et al., 2015). This prediction is supported by previous findings on multimodal emotional convergence that suggest presentation of congruent bimodal emotional cues (e.g., verbal and nonverbal) yield faster and more accurate emotion judgments than unimodal presentations (e.g., only verbal) (Calvert, 2001; Schirmer & Adolphs, 2017; Schröger & Widmann, 1998).

To test this hypothesis, we focused on the affective dimension of arousal and organized words in two groups of iconic and non-iconic by the orthogonal manipulation of the factors *lexical arousal* and *sublexical arousal* (Figure 8.1). In a two-alternative (high arousing vs. low arousing) forced choice task, we then asked participants to decide as quickly and accurately as possible whether the meaning of visually presented words was "exciting" or "calming" (i.e., an arousal decision task). Note that at both *lexical* and *sublexial* levels our experimental design involves primarily the manipulation of arousal rather than valence. At the *sublexical* level, arousal plays a dominant role in models of vocal emotion communication (Bachorowski, 1999; Bänziger, Hosoya, & Scherer, 2015) and in shaping affectivity in a word's sound (Aryani, Conrad, et al., 2018). At the *lexical* level, the first emotional appraisal of a stimulus has shown to be related to arousal which qualifies it as the primary factor producing emotional interference in information processing tasks (Anderson, 2005; Dresler, Mériau, Heekeren, & Van Der Meer, 2009; Schimmack, 2005). Thus, with regard to rather faster sensory processing of words' sound, arousal seems to be a more suitable candidate for an interactive effect between sound and meaning. Note that since the decision response time for a forced choice task had to be measured accurately, words in this study were presented visually. Therefore, it is important to mention that the use of the term "sound" in the present work refers to the implicit sound of words derived from phonological and prosodic recoding (Braun et al., 2009; Breen, 2014; Ziegler & Jacobs, 1995).

8.3 Materials and Methods

The study was approved by the ethics committee of the Freie Universität Berlin and was conducted in compliance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). All participants gave their consent in written form prior to participating in the study.

8.3.1 Stimuli

One hundred and sixty nouns (one to three syllables long) were selected for a 2×2 design involving twofold manipulations of *lexical* and *sublexical arousal* (see Figure 8.1). For *lexical arousal*, we used ratings for words' affective meaning (min = 1: very low arousing, max = 5 very high arousing) from the normative database BAWL-R (Võ et al., 2009). *Sublexical arousal* was calculated using the recent psychoacoustic model (Aryani, Conrad, et al., 2018). This model is based on specific extracted acoustic features of pseudowords (e.g., pitch, formants, and intensity) that predict ratings given on the affectivity of their sound (see study2b in Aryani, Conrad, et al., 2018).

Words were then divided into two distinctive conditions of "high" and "low" arousing for each of factors: *lexical arousal* ("high" > 3.25, "low < 2.75) and *sublexical arousal* ("high" > 3, "low" < 3) and carefully matched for relevant psycholinguistic factors (see Table 8.1).

Due to a natural confound between affective arousal and valence, words in the condition of "high" *lexical arousal* were more negative in valence than words in the condition “low” *lexical arousal*. In order to prevent participants to build an alternative strategy basing their decision on valence rather than arousal, 60 filler words with the rather rare combination of high *lexical arousal* and positive *lexical valence*, as well as 60 words with low *lexical arousal* and negative *lexical valence* were added to the stimulus set, which were excluded from further analyses.

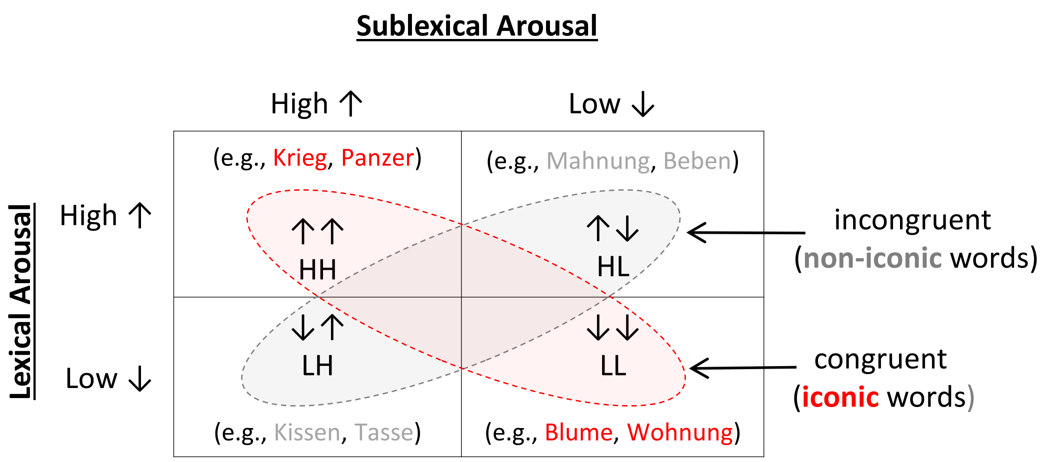


Figure 8.1. Words were organized in a 2 × 2 design with each of experimental factors (*lexical arousal* and *sublexical arousal*) manipulated in two distinct groups consisting of extreme levels of arousal (High = exciting, and Low = calming). The congruence vs. incongruence of *lexical arousal* (meaning) and *sublexical arousal* (sound) resulted in two groups of iconic vs. non-iconic words, respectively. Two example words (in German) from each category are given in each cell.

8.3.2 Participants

Thirty-six right-handed German native speakers (26 women, mean age: 22.5 years, range: 18–34 years) with no history of neurological or psychiatric illness volunteered to participate in the study, receiving either five Euros or psychology course credit for their participation. All participants reported normal or corrected-to-normal vision and provided written informed consent to participate in the study. Handedness was determined using the Edinburgh Inventory (Oldfield, 1971).

Variable	Word Category								Inferential Statistics
	HH		HL		LH		LL		
	M	SD	M	SD	M	SD	M	SD	
Lexical Arousal	3.86	0.43	3.75	0.35	2.13	0.28	2.17	0.32	$F(3,156) = 301, p < 0.0001$
Lexical Valence	−1.59	0.66	−1.44	0.63	1.03	0.67	1.04	0.74	$F(3,156) = 190, p < 0.0001$
Sublexical Arousal	3.00	0.25	2.17	0.15	2.97	0.22	2.17	0.14	$F(3,156) = 230, p < 0.0001$
Word Frequency	0.97	0.67	0.88	0.75	0.90	0.74	0.93	0.66	$F(3,156) = 0.11, p = 0.95$
Imageability Rating	4.44	1.27	4.31	1.08	4.32	1.43	4.27	1.37	$F(3,156) = 0.12, p = 0.94$
# Syllables	2.13	0.56	2.08	0.41	2.15	0.57	2.08	0.47	$F(3,156 = 0.21, p = 0.88$
# Letters	6.05	1.22	6.05	1.20	6.08	1.40	6.00	1.30	$F(3,156) = 0.02, p = 0.99$
# Phonemes	5.53	1.20	5.33	1.02	5.45	1.22	5.20	1.07	$F(3,156) = 0.64, p = 0.59$
# Orth-Neighbors	1.40	1.69	1.08	1.79	1.45	2.00	1.75	2.06	$F(3,156) = 0.85, p = 0.46$
Orth-Neighbors-HF	0.50	0.91	0.43	1.26	0.48	0.93	0.50	0.99	$F(3,156) = 0.04, p = 0.98$
Orth-Neighbors-Sum-F	0.72	1.06	0.49	0.82	0.69	1.05	0.80	0.88	$F(3,156) = 0.76, p = 0.51$
# Phon-Neighbors	1.75	2.51	1.98	3.04	1.93	2.58	2.35	3.34	$F(3,156) = 0.30, p = 0.82$
Phon-Neighbors-HF	0.55	0.88	0.63	1.76	0.55	1.08	0.60	1.24	$F(3,156) = 0.03, p = 0.99$
Phon-Neighbors-Sum-F	0.79	1.02	0.69	1.01	0.67	0.92	0.88	0.88	$F(3,156) = 0.40, p = 0.75$

Table 8.1. Characteristics of word stimuli

8.3.3 Procedure

Participants were instructed to decide, as quickly and correctly as possible, whether the meaning of a word presented visually was either high or low arousing (exciting or calming), and to correspondingly press one of two designated buttons on the keyboard (in German: “*Deine Aufgabe ist es, so schnell und so korrekt wie möglich zu entscheiden, ob du die Bedeutung des präsentierten Wortes als aufregend oder beruhigend empfindest….Für deine Entscheidung verwende bitte die beiden Tasten (...) für aufregend und (...) für beruhigend*”). The assignment of the response buttons was counterbalanced across participants. Participants worked through 10 practice trials before starting with the 280 (160 experimental + 120 distractors) main trials. Each trial started with a fixation cross in the screen center with a jittered duration between 1.5 and 3 s and continued with the stimulus item being presented for 1.5 s or until a decision was made. The order of item presentation was fully randomized. For each item, we recorded the response of the first button press.

After the decision task, in a separate study, the same participants were asked to rate the same 160 relevant words for *lexical arousal*. Adapting the instructions used for the original BAWL ratings (Võ et al., 2009), participants were invited to read the presented item and evaluate how exciting or calming the presented word means. The 5-point affective sound of arousal scale ranged from 1 (*sehr beruhigend*/"very calming") to 5 (*sehr aufregend*/"very exciting"). We also incorporated the self-assessment manikins (SAM) that were used in the ANEW study (Bradley & Lang, 1999). The items were randomly presented to minimize primacy or recency effects. We then used these rating values as a reference for evaluating responses given in the decision task, thereby distinguishing between "wrong" responses and "subjectively different" responses.

8.3.4 Analysis

Trials without response were excluded from the analyses (2%, N = 110). We then compared the responses of each participant with their own affective judgment given in the rating study. Responses in the decision task that were in accordance with the rating values, but not in alignment with the original ratings used in experimental manipulation, or vice versa, i.e., *subjectively different* responses, were excluded from the analyses (17%, N = 1002), leaving 447 *wrong* responses (7%) and 4201 *correct* responses (73%). Using language stimuli, we chose Linear Mixed Model (LMM) analysis—over the classic F_1-F_2 test—which provides a solution for the long-standing problem of how to analyze experimental data that contain two crossed random effects, i.e., items and participants (see for instance (Janssen, 2012) for a review). RT and accuracy data for the items were analyzed with a linear mixed fixed and random effects model using the statistical software JMP 13Pro (SAS Institute Inc.), with *lexical* and *sublexical arousal* and their interaction as fixed effects and participants and items as random effects.

In order to ensure that the exclusion of a large amount of responses (none and *subjectively different*) was randomly distributed across experimental conditions and did not bias the results, we took the 1112 excluded words and ran the same mixed model analysis predicting the RT within these excluded items.

8.3.5 Results

A comparison between original ratings for *lexical arousal* (from the BAWL) and the average of post hoc ratings revealed a high consistence between values: r = 0.94, $p < 0.0001$, indicating the reliability of the used measure for *lexical arousal* as experimental factor.

The analysis of the excluded responses showed that the distribution of these items across experimental conditions was very similar over congruent (9.8%, N = 568) vs. incongruent conditions (9.4%, N = 544) and not significantly different over participants ($p = 0.96$). Within the excluded items, there was no significant effect of any of the experimental factors on the reaction time nor a significant interaction (all ps > 0.3), suggesting that the exclusion of items did not follow a systematic pattern, and consequently, did not bias the results of the remaining responses.

Results of two main LMM analyses on remaining responses are displayed in Figure 8.2 and Table 8.2. A significant effect of *lexical arousal* on accuracy and on RT was observed with lexically high-arousing words classified more correctly and more quickly than low-arousing words (both ps < 0.001). No direct effect of *sublexical arousal* on response accuracy or on RT was observed ($p = 0.57$, $p = 0.48$, respectively). Importantly, there was a significant interaction between *lexical* and *sublexical arousal* for both accuracy and RT (both ps < 0.05). Post hoc analysis showed that within each *lexical* category, iconic words were associated with a higher response accuracy and a shorter RT than non-iconic words (see Table 8.2 for further results).

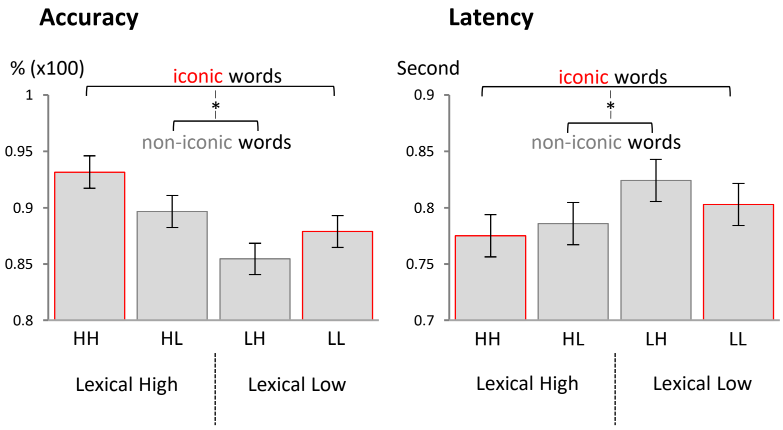


Figure 8.2. Congruent words (iconic) were classified more quickly (right) and more accurately (left) in the corresponding lexical group compared to incongruent words (non-iconic).

Response Accuracy					Response Latency				
Term	Estimate	Std E	*t*	*p*	Term	Estimate	Std E	*t*	*p*
Intercept	0.903	0.010	90.07	<0.0001	Intercept	0.797	0.017	45.47	<0.0001
lexical arousal	0.018	0.005	3.55	0.0005	*lexical arousal*	−0.016	0.003	−4.32	<0.0001
sublexical arousal	0.002	0.005	0.56	0.5784	*sublexical arousal*	0.002	0.003	0.69	0.4899
*lexical*sublexcial*	0.013	0.005	2.55	0.0120	*lexical*sublexcial*	−0.008	0.003	−2.11	0.0369

Table 8.2. Results of fixed effects, the interaction term, and the intercept of the mixed model analysis.

8.4 Discussion

In this study, we investigated the effect of iconicity on affective semantic decisions and tested whether language users take the sound aspect implicitly into consideration. In line with our H1, faster latencies and higher accuracy in responses were observed for iconic words, i.e., words that exhibit similarity between meaning and sound in affective domain. Our finding, thus, clearly shows that in the context of language processing, human subjects are sensitive to affective cues that are provided by words' sound even when they are presented visually and read silently. Such affective cues can be integrated in higher cognitive processes and affect semantic decisions, thereby facilitating the evaluation of words’ affective content when sound and meaning aspects are congruent. Crucially, this effect is evident even when the attentional focus is not directly on the sound aspect of words, suggesting an implicit effect of sound on the evaluation of words’ meaning (see also Aryani, Conrad, et al., 2018; Schirmer & Kotz, 2006). With this study, we aimed to build upon the previous results on the facilitatory effect of iconicity in lexico-semantic processing, which has been reported in sign language (Thompson et al., 2012; Vinson et al., 2015b), in onomatopoetic words and in ideophones (Dingemanse et al., 2016; Kita, 1997; Kwon & Round, 2014; Lockwood et al., 2016). By using quantitative measures for both sound and meaning of words, we extended the results of previous findings to a larger number of “ordinary” words in the lexicon and in the context of affective meaning.

Also, the important role of multimodal convergence of emotions in making appropriate and faster decisions in emotional evaluation is supported by our data. A major benefit of multimodal integration has been shown to optimize efficient information processing by minimizing the uncertainty of ambiguous stimuli (see (Klasen, Chen, & Mathiak, 2012; Schirmer & Adolphs, 2017) for recent reviews). This is well in line with our behavioral results, in which words possessing congruent affective information from two different sources (i.e., sound and meaning) were categorized more quickly and more accurately.

The observed effect of lexical arousal on latency and accuracy also supports the previous findings on preferential processing of high arousing compared to low arousing words in decision tasks (e.g., Hofmann, Kuchinke, Tamm, Võ, & Jacobs, 2009), which is proposed to be rooted in a biologically adaptive response leading to a faster and more accurate evaluation of emotionally relevant stimuli.

Importantly, in line with the results of previous investigations (Aryani, Conrad, et al., 2018; Westbury, 2005), the effect of iconicity facilitates lexico-semantic processing of words even when they are visually presented and silently read (see (Ullrich et al., 2016) for an ERP study for the effect of implicit sound). Note that visual word recognition generally involves the activation of phonological codes (Braun et al., 2009; Breen, 2014; Ziegler & Jacobs, 1995) and language users appear implicitly influenced by affective sound of visually presented words when evaluating the affective meaning of these words (Aryani, Conrad, et al., 2018). However, as we did not control our stimuli for orthographic features, a possible effect of graphemes on the processing of the affective content of words (Cannon, Hayes, & Tipper, 2010; den Bergh, Vrana, & Eelen, 1990) is not precluded.

With the present study, we also aimed at drawing attention to the role of emotion in language processing, and in particular, in the study of iconicity. Focusing only on perceptuomotor analogies between sound and meaning, previous studies often overlooked investigating emotion as a modality of experience similar to sensory and motor processing (Aryani, Conrad, et al., 2018; Jacobs, Hofmann, et al., 2016; Vigliocco et al., 2009). Affective meaning is, however, a fundamental aspect of human communication that have been proposed as the original impetus for language evolution (Darwin, 1888; Panksepp, 2010). Therefore, from a phylogenetic perspective, the effect of iconicity may be most evident in the affective communication. Here, iconicity serves as an interface for accomplishing the need to map linguistic form onto human affective experience as a vital part of meaning making. When analysing the results, we had to exclude a relatively large number of items (17%) that were differently rated from the original ratings used in the experimental manipulation (i.e., the BAWL ratings). This may call for cautious interpretation of the results as it raises a question about the nature of arousal as a semantic feature. A more detailed analysis of these items did not reveal any specific pattern in regard to the degree of arousal nor to a specific group of words. Previous rating studies have repeatedly shown that ratings of valence are relatively consistent across participants while arousal is much more variable (Stadthagen-Gonzalez, Imbault, Pérez Sánchez, & Brysbaert, 2017; Võ et al., 2009; Warriner, Kuperman, & Brysbaert, 2013). It has been suggested that valence is a semantic super-feature that results from an integration of both experiential and distributional data (Jacobs, Hofmann, et al., 2016) as assumed by the semantics theory of Andrews et al. (Andrews, Vigliocco, & Vinson, 2009). Arousal, however, may be derived by way of experience with the physical world and thus being less distributional (i.e., language based) and more experiential (i.e., non-language based). This, in turn, can explain the individual differences of arousal ratings at the level of meaning and, at the same time, its consistence at the level of sound leading to its dominant role in models of vocal emotion communication (Bachorowski, 1999; Bänziger et al., 2015) as outlined in the introduction.

Concerning the nature of sound-meaning mapping, two different types of mapping, i.e., iconicity and systematicity, have been suggested in the previous work (Dingemanse et al., 2015). The sound-meaning mapping in a word is considered iconic when both sound and meaning independently refer to a similar specific (sensory, motor, or affective) domain (Perniss & Vigliocco, 2014). For instance, some swear words are considered iconic because both their sound and their meaning possess negative valence (Bowers & Pleydell-Pearce, 2011). In the present study, we used two different measures for assessing the sound and meaning of words based on their affective arousal. At the meaning level, our measure for the lexical arousal has been cross-validated in various empirical studies regarding experiential, behavioral, and neurobiological levels of analysis (Jacobs et al., 2015). Also, at the sound level, the measure of sublexical arousal used in this study has been shown to have an inherent affective quality based on acoustic features that are known to modulate nonverbal emotional communication (Aryani, Conrad, et al., 2018) and can evoke affective brain responses similar to other types of affective sounds (Aryani, Hsu, et al., 2018). Consequently, it is reasonable to conclude that our finding on the facilitatory effect of sound-meaning mapping is related to iconic mappings of words rather than statistical regularities in the lexicon.

Our finding can also help to gain a better understanding of affective and aesthetic processes of literary reading (Jacobs, 2015a; Schrott & Jacobs, 2011). Poetry, for instance, seems to be one of the most promising forms of literature for sound-meaning investigations. The relation of "form" to "feeling" supposedly lies at the basis of poetry (Langer, 1953), and the "*differentia specifica*" of poetry is located in its formal characteristics and iconic properties (Jakobson, 1960a). Poetry is on the one hand inherently concerned with emotional expressions, and on the other hand, is accompanied by the artful deployment of sound patterns (Aryani et al., 2013, 2016; Jacobs, Lüdtke, et al., 2016; Lüdtke et al., 2014; Schrott & Jacobs, 2011; Ullrich et al., 2017). In this context, our results on the facilitated lexical processing of iconic words can be linked to previous findings on the notion of processing fluency stating higher ease of processing leads to a higher aesthetic pleasure (Bohrn et al., 2012a; Reber et al., 2004). This may provide additional explanation for the preferential use and the aesthetic effects of stylistic devices such as phonaesthetics and iconicity in poetry.

Chapter 9

Affective Iconic Words Benefit from Additional Sound-Meaning Integration in the Left Amygdala[5]

9.1 Abstract

Recent studies have shown that a similarity between sound and meaning of words (i.e., iconicity) can facilitate semantic decisions. However, the neural mechanisms underlying the beneficial role of iconicity are largely unknown. By focusing on the affective domain, we tested the hypothesis that affective iconic words benefit from an additional processing network consisting of neural hubs which integrate affective information from different sources, i.e., *sublexical arousal* related to acoustic information and *lexical arousal* related to semantic information. Iconic words, compared to their non-iconic counterparts, were associated with increased fMRI signals in the left amygdala known for its role in multimodal representation of emotions. Results of functional connectivity analyses demonstrated that the observed amygdalar activity is modulated by activations in the left superior temporal gyrus and the left inferior frontal gyrus, representing processing hubs of sound and meaning, respectively. Our data indicate the involvement of an additional processing network for iconic words based on multimodal integration. These results suggest that language

[5] This chapter has been submitted for publication as: Aryani, A., Hsu, C. T., & Jacobs, A. M. (Submitted). Affective Iconic Words Benefit from Additional Sound-Meaning Integration in the Left Amygdala

users are sensitive to iconic mappings and rely on affective cues in the sound of words to evaluate the affective meaning.

9.2 Introduction

The evolutionary jump from the use of inherent and motivated signs (e.g., a cave painting of a horse representing a horse) toward building unmotivated and arbitrary signs (e.g., using the word 'horse' to represent a horse), has been suggested to lay the groundwork for why humans have language (e.g., Deacon, 1997). Thus, arbitrariness of linguistic sign is considered one of the most fundamental properties that grants human language its compositional power, referential flexibility, and productivity (Saussure, 1916/2011; Gasser, 2004; Hockett, 1958; Monaghan et al., 2011), setting humans apart in the animal kingdom by means of a remarkably unique communication system.

However, in contrast to the notion of the absolute arbitrariness of the linguistic sign, recent empirical data suggest a stand-alone role of sound in meaning making beyond arbitrary and conventional links (Dingemanse et al., 2015; Perniss, Thompson, & Vigliocco, 2010b; Schmidtke, Conrad, et al., 2014). A prominent type of such non-arbitrary mapping between sound and meaning is iconicity in which sound imitates or resembles some aspects of the meaning. Onomatopoeia (e.g., 'click') and ideophones (e.g., 'zigzag') represent most instructive examples for this category of words that can evoke sensory, motor, or affective experiences directly or indirectly, respectively (Schrott & Jacobs, 2011).

Iconic words have been suggested to be capable of directly activating the semantic domain that they refer to by bridging the gap between linguistic form and human (sensory, motor and affective) experience (Aryani & Jacobs, 2018; Perniss & Vigliocco, 2014; Vinson et al., 2015). Thus, iconicity may provide additional mechanisms for both vocabulary learning and language processing by means of direct sound-meaning mappings in neural systems devoted to perception, action and affective experience; a mechanism that can potentially realize the embodiment of language (Aryani, Conrad, et al., 2018; Meteyard et al., 2015a; Vigliocco, Perniss, & Vinson, 2014; Vinson et al., 2015a).

A small number of studies provided the first neural evidence indicating that iconic words profit from an additional sublexical (i.e., sound) processing network. For instance, Kanero et al. (2014) compared onomatopoetic expressions that were related to motion and shape with arbitrary words from the same semantic domains. Results showed greater general activation, and a cluster of activation in the right posterior superior temporal sulcus (pSTS) presumably working as a hub for integration of multimodal (i.e., lexical and sublexical) information. This finding aligns with the results of previous work on onomatopoetic words showing a greater activation for bimodal information (i.e., onomatopoetic words imitating animal calls) in the left and right superior temporal sulcus (STS) than for unimodal information (i.e., either animal names, or animal calls; (Hashimoto et al., 2006). By extending the word material to a multi-language stimulus set, Revill et al. (2014) provided further support for the advantageous processing of iconic words, and for the potential role of areas engaged in multimodal sensory integration beyond those involved in semantic processing. These results suggest the existence of more direct links between semantic information and sound information for iconic words with corresponding neural hubs as convergence zones for information integration. Iconic words might therefore be more immune to neurological damages that affect language processing networks as, for instance, in aphasic patients. In fact, in a recent lesion study involving individuals with aphasia following left-hemisphere stroke (Meteyard et al., 2015), a consistent processing advantage was observed for onomatopoetic words in reading aloud and auditory lexical decision; two tasks that rely on sound-meaning mapping.

Overall, previous studies suggest that iconicity can facilitate language processing through activation of additional links between the sound of a word and modality-specific experiences (i.e., sensory, motor, affective), as well as through integration of information from different modalities which may provide an opportunity for stronger embodiment of iconic signs (Aryani, Conrad, et al., 2018; Vigliocco et al., 2014). Nevertheless, unlike pioneering work on the facilitatory effect of iconicity in sign language (Thompson et al., 2012; Vinson et al., 2015) which laid the ground for the theoretical framework of such investigations, related research on spoken language—including behavioral studies—has so far mainly focused on marginal cases in language practice, i.e., onomatopoeia and ideophones including Japanese mimetic words (Dingemanse et al., 2016; Iwasaki et al., 2007a, 2007b; Kwon & Round, 2014; Lockwood et al., 2016), or on cases typically not considered iconicity but rather statistical regularities in vocabulary (i.e., systematicity; Bergen, 2004; Farmer et al., 2006; Reilly et al., 2012).

For the present study, we built upon the results of two recent studies that capitalize on the affective domain showing an influence of the implicit sound of words on the evaluation of affective meaning (Aryani, Conrad, et al., 2018), as well as a facilitatory role of iconicity in 'ordinary' words (Aryani & Jacobs, 2018). Based on the congruence vs. incongruence of lexical (meaning) and sublexical (sound) arousal, words in the latter study were organized in two groups of iconic vs. non-iconic, and presented in a semantic decision task. Results showed a beneficial processing of iconic words, compared to their non-iconic counterparts, in terms of shorter latencies and higher accuracy of decision responses. This finding suggests that affective cues in the sound of a word (implicit or explicit) can be integrated with higher-order semantic processes facilitating the evaluation of affective content when sound and meaning aspects are congruent.

Using event-related fMRI in the present study we aimed at exploring the neural mechanisms underlying such beneficial processing of affective iconic words. For this, we used a similar experimental design as in the aforementioned study, characterized by an orthogonal manipulation of *lexical* and *sublexical arousal* (representing meaning and sound, respectively, see Figure1) and presented words in a passive listening task (see Methods for details). We predicted a generally greater activation for iconic words than non-iconic words, particularly in brain areas associated with affective processing, and in convergence zones responsible for multimodal representation of emotional information from different sources: i.e., acoustic information (related to *sublexical arousal*) and semantic information (related to *lexical arousal*).

Candidates for regions integrating emotional information from different domains have been found in previous work by focusing mostly on the integration of audiovisual cues, as well as verbal and nonverbal vocal cues. These extend from higher association areas such as the anterior and the posterior cingulate cortex (ACC and PCC), to prefrontal cortex (PFC), and (left) amygdala (Chen & Spence, 2010; Klasen, Kenworthy, Mathiak, Kircher, & Mathiak, 2011; Klasen et al., 2012; Wittfoth et al., 2009). However, among these areas, PFC and ACC responded more strongly to incongruent than congruent emotional information in accordance with their prominent role in the conflict network (Botvinick, Cohen, & Carter, 2004; Etkin, Egner, & Kalisch, 2011; Hofmann, Tamm, et al., 2008; Kerns et al., 2004). Therefore, we expected to observe activation in PCC and/or the left amygdala—as supramodal emotion integration networks— for iconic words due to the congruence between sublexical and lexical affective information.

To investigate the interaction between brain regions involved in iconic sound-meaning mappings, we also performed a functional connectivity analysis and adopted two independent seed regions representative for processing of sound (left superior temporal gyrus, STG) and meaning (inferior frontal gyrus, IFG) of words (see Methods). We hypothesized that iconicity significantly increases the coupling between these two seeds, on the one side, and the convergence zones integrating emotional information (PCC and/or the left amygdala), on the other side.

At the end of the experiment, outside the scanner, an unannounced recognition test was performed to assess participants' involvement in the task and mnemonic effects of the experiment. After the recognition test, in order to check the reliability of our experimental manipulations participants were asked to evaluate the words presented in the scanner for their *lexical* and *sublexical arousal* in two separate rating studies.

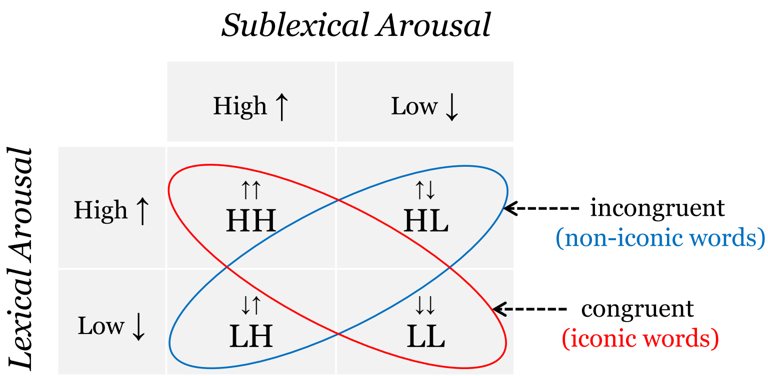


Figure 9.1. Word stimuli were organized in a 2x2 design: with each experimental factor (*lexical* and *sublexical arousal*) manipulated in two distinctive groups consisting of extreme levels of arousal (High=exciting, and Low=calming). The congruence vs. incongruence of *lexical* (meaning) and *sublexical arousal* (sound) results in two groups of iconic vs. non-iconic words, respectively.

9.3 Results

9.3.1 Behavioral Results

Recognition Test: Across all participants, we performed a linear mixed model (LMM) analysis predicting the recognition rate, with word category (OLD = 'words used in the scanner' vs. NEW = 'words not used in the scanner') as fixed factor and words as well as participants as random factors. Results supported a performance above chance for recognizing OLD words, with a significantly higher score (M=3.53) compared to NEW words ($M = 2.54$, $t = -20.6$, $p < .0001$). We next performed simple t-tests to compare the recognition rate between the levels of word category (OLD vs. NEW) separately for each participant. An effect of word category (OLD vs. NEW) on accuracy was observed for 27 participants out of 29 ($t = 6.4 \pm 3.2$). These results indicate that the majority of participants had been attentive during the passive listening task. Two participants with a performance not higher than chance level ($t = 0.28$, $t = 1.14$) were consequently excluded from further analyses.

Ratings: *Lexical* and *sublexical arousal* ratings used for stimulus construction were correlated with our post-scan data. For both, the correlation coefficients were very high: $r = .97$, $p < .0001$, (r_{min} among all participants = .73), and $r = .76$, $p < .0001$ (r_{min} among all participants = 0.49), respectively.

9.3.2 Neuroimaging Results

GLM Results:

The comparison between all words contrasted with the auditory baseline condition (SCN, see Method) revealed left-lateralized activations in core language areas, i.e., the inferior frontal gyrus (IFG), middle and superior temporal gyrus, and inferior parietal lobule (BA 40), suggesting that this experiment successfully tapped into the language processing system. Activity was also observed in bilateral parahippocampal gyrus, middle frontal gyrus and precentral gyrus, as well as the left superior frontal gyrus, the fusiform area, the right caudate, and superior parietal lobule.

Results of the two main effects of *lexical arousal* (Lex H > Lex L) and *sublexical arousal* (Sub H > Sub L) were reported and discussed in detail elsewhere (Aryani, Hsu, Jacobs, 2018). In summary, the comparison Lex H > Lex L was associated with activation in brain regions involved in appraisal and general processing of affective stimuli, i.e., dorsolateral and medial prefrontal cortex, posterior cingulate cortex, the left inferior frontal gyrus, and temporal pole. Brain regions associated with the comparison Sub H > Sub L were substantially similar to those involved in processing of other types of affective sounds, i.e., superior temporal area, bilateral insula, and premotor cortex (Aryani, Hsu, Jacobs, 2018). These results, together with those of the comparison Words > SCN, as well as of the behavioral studies strongly support the reliability of our experimental manipulations and show that this experiment successfully engaged participants in carefully listening to words.

The present contrast between congruent and incongruent words regarding *lexical* and *sublexical arousal* (Iconic > Non-Iconic) revealed a cluster of significant activation in the left amygdala with an activation peak in [x y z] = [-29 -6 -14], and cluster size of $k = 574$ voxels (Figure 2). Results of a repeated analysis with the reduced voxel-level CDT of 0.001 (see methods) showed, again, a similar significant cluster (FWE-corrected p-value < 0.05) in the left amygdala with a smaller size (k = 279), and with a peak activation at the same location as in the former analysis. No cluster of activation was found for the reverse contrast (Incong>Cong).

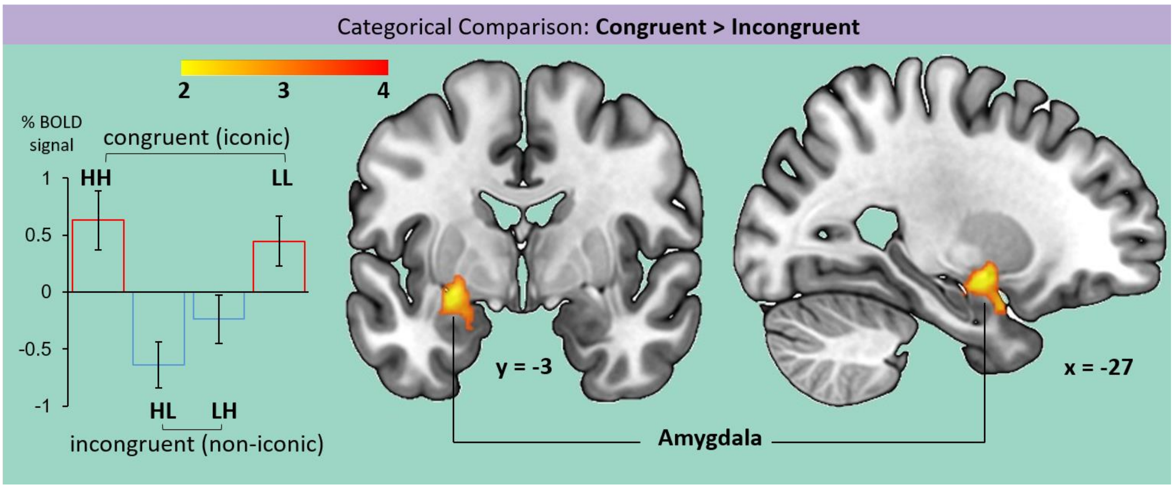


Figure 9.2. Iconic words as defined by the congruence between *lexical* and *sublexical arousal* elicited BOLD signals in the left amygdala (p<0.05, FEW-corr). Pairwise comparisons showed increased activation in the same region for the contrast HH>HL, as well as LL>LH.

<u>PPI Results:</u>

At the whole brain level, PPI analysis did not reveal any functional connectivity in the iconic condition (Cong > Incong). Using small volume correction, we defined a region of interest (ROI) based on the anatomical amygdala map using the WFU Pickatlas Tool. The ROI analysis revealed a significant cluster in the left amygdala (activation peak [x y z] = [-29 -6 -14], and cluster size of k = 157 (Figure 3).

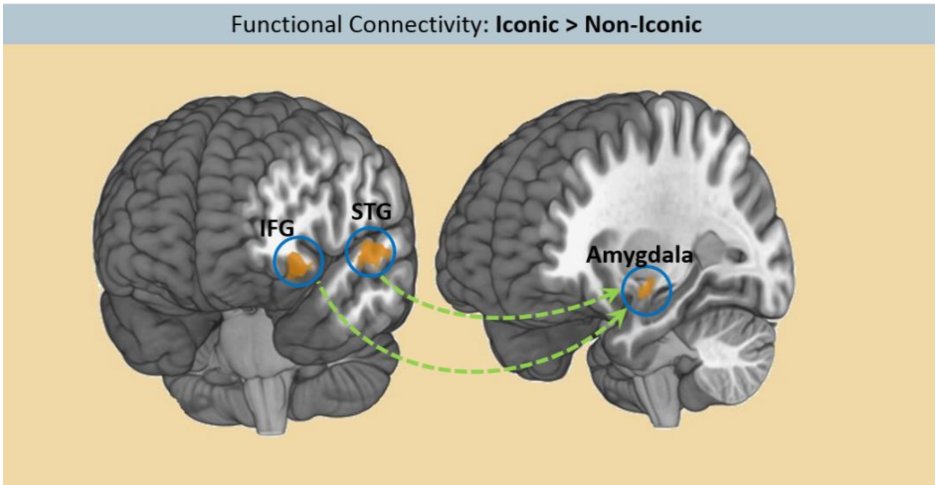


Figure 9.3. In the congruent condition (iconicity), left amygdala showed significant functional connectivity with activation in two seed regions: the left superior temporal gyrus (STG) and the left inferior frontal gyrus (IFG) representing the processing of sound and meaning of words, respectively.

9.4 Discussion

The present study examined to what extent iconic words—as defined by the congruence between affective sound and affective meaning—profit from an additional processing network that integrates the affective information in the sound and meaning of words. In line with our hypothesis an interaction between affective information from two different sources (i.e., words' sublexical affective sound and lexico-semantic affective meaning) was observed as reflected in the left amygdala activity. Also, pairwise comparisons showed increased activation in the same region within two groups of lexically high and low arousing words (HH>HL, and LL>LH, see Figure 2). In addition, our functional connectivity analysis demonstrated that the observed activity in the left amygdala is modulated by activation in the left STG and the left IFG; two brain regions known for their prominent roles in sound and meaning processing.

The activation of the left amygdala in response to congruent emotional information from the sound and meaning of words is in line with its proposed role in supramodal emotion integration, and functioning as a general convergence zone (Klasen et al., 2011; Kreifelts, Ethofer, Huberle, Grodd, & Wildgruber, 2010; Müller, Cieslik, Turetsky, & Eickhoff, 2012; Schiller, Freeman, Mitchell, Uleman, & Phelps, 2009). The amygdala has reciprocal connections with association cortices in the superior and inferior temporal gyri (Aggleton, 1993) through which it can selectively modulate sensory responses depending on their emotional relevance. Thus, the amygdala appears to act as a neural gateway for binding the information from different modalities with each other, and also with brain region associated with emotional and motivational information. This view is in line with the modulatory role of amygdala in a wide array of networks and its functional importance in broader and more abstract dimensions of information processing (Jacobs, Hofmann, et al., 2016; Pessoa & Adolphs, 2010).

Our results share a substantial similarity to the results of previous work on neural mechanisms of affective prosody suggesting a network involving interactions between the STG and the IFG (Ethofer et al., 2011; Frühholz & Grandjean, 2013; Glasser & Rilling, 2008; Leitman et al., 2016) with the STG forming emotional representations of acoustic features (Aryani, Hsu, Jacobs, 2018; Frühholz, Ceravolo, & Grandjean, 2011; Leitman et al., 2016; Wiethoff et al., 2008), and the IFG evaluating the meaning and the relevance of the sound (Leitman et al., 2010; Schirmer & Kotz, 2006; Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006). Importantly, these studies, in line with our results, suggest that final appraisal of affective prosody takes place in the amygdala (Frühholz et al., 2016; Leitman et al., 2016) when explicitly listening to the voice, as well as, when emotional voices are presented outside the current focus of attention (Frühholz et al., 2011; Frühholz & Grandjean, 2013; Leitman et al., 2016).

The fact that we did not observe significant activation in the PCC, as expected in the introduction, might be due to the lack of socially relevant information in our stimuli. Previous studies on bimodal emotion integration mostly used human faces and voices both of which rely on social information potentially explaining the activation in the vPCC for emotionally congruent stimuli (Klasen et al., 2011; Schiller et al., 2009) However, unlike human faces and voices, the affective sound of words is based on basic acoustic features (Aryani, Conrad, et al., 2018; Myers-Schulz et al., 2013) and is processed in substantially similar brain networks as other types of non-human affective sounds (Aryani, Hsu, Jacobs, 2018).

Interestingly, unlike the results of multimodal integration of incongruent emotions, the inverse contrast for non-iconic words in our study (incong > cong) did not elicit any significant cluster of activation. Although a neural effect of incongruent stimuli could intuitively be anticipated in brain regions associated with the conflict network, i.e., PFC and ACC (Botvinick et al., 2004; Etkin et al., 2011; Hofmann, Herrmann, et al., 2008; Kerns et al., 2004), the lack of significant activation for this contrast suggests that the human brain does not treat arbitrary relationships between sound and meaning as conflict. In fact, as the majority of words in the language are learned through conventional and per se arbitrary links, these results suggest that non-iconic words are chiefly processed in the core language regions, and even in the case of incongruence between sound and meaning no extra processing is devoted. Future research is needed to examine in more detail the neural substrates of incongruent words vs. a neutral baseline by using experimental designs with more distinct levels for *lexical* and *sublexical arousal* (e.g., high, medium, and low).

A particular aspect of this study was that the selection of words went beyond the previous focus on marginal group of iconic words in language (i.e., onomatopoeia and ideophones), thereby providing novel evidence on the effect of iconicity in language processing as a general property of lexicon. A crucial issue of the use of onomatopoetic words in previous investigations on iconicity relates to the limitation of implicit testing of the sound effect. Usually possessing a special phonological construction, this type of words can possibly raise an undesirable awareness to the sound aspect of words when using in an experimental setting. Moreover, previous work reported a strong variation in the distribution of onomatopoeia and ideophones in different languages: while, for instance, Indoeuropean languages are considered as less iconic, sub-Saharan African and Southeast Asian languages seem to be richer in the use of such iconic mappings (Perniss et al., 2010; Watson, 2001). This can potentially hinder the cross linguistic investigations in this line of research and prevent the possible replication studies when focusing only on this type of iconic words (i.e., onomatopoeia). By investigating the effect of iconicity in ordinary words, we addressed this limitation apparent in previous studies and provided neural evidence for iconicity as a general feature of language practice (see Perniss et al., 2010).

In terms of language processing, iconicity provides an additional mechanism for the grounding of language in sensory, motor, and affective systems. Iconic mappings imply the engagement of such systems by means of their very nature of depicting the corresponding properties of what they refer to; i.e., sensory, motor, or affective experience. Consequently, iconicity should have facilitative effects in language processing as it would render the link between form and meaning stronger (see also Aryani, Hsu, et al., 2018; Meteyard et al., 2015). Our data firmly support the existence of such a direct link between the sound and meaning for affective iconic words. This sheds new light to mechanisms underpinning the advantages of iconicity for language processing and learning vocabulary (Nygaard et al., 2009).

Showing a greater engagement of affective brain regions for (affective) iconic words, our finding can advance the understanding of affective and aesthetic processes of literary reading (Jacobs, 2015a; Schrott & Jacobs, 2011). In line with its role in multimodal emotion integration, the left amygdala has been proposed to respond to metaphoric language, valence congruity, figurativeness, and harmony (Jacobs, Hofmann, et al., 2016). Empirical support for this view comes from studies showing an enhanced left amygdala activation for metaphors (Citron & Goldberg, 2014) and metaphorical Noun-Noun-Compounds (Forgács et al., 2012) when compared to their literal counterparts. Also, results of a meta-analysis of 23 neuroimaging studies showed a left amygdala activation in response to a variety of figurative statements, and in particular metaphors (Bohrn, Altmann, & Jacobs, 2012a). The meaning of metaphors is in general based on considerations of similarity between different aspects of target and source, and this is what iconicity in language is about. Lakoff and Turner (Lakoff & Turner, 1989) defined iconicity as a “metaphorical image-mapping in which the structure of the meaning is understood in terms of the structure of the form of the language presenting the meaning". Such image-mapping, according to them, is enabled by image-schemas which are formed from our embodied experience. This view emphasized the role of the left amygdala as a central hub critical for regulating the flow and integration of information from different experiences.

9.5 Conclusion

The present data indicate that language users are sensitive to the interaction between sound and meaning aspect of words, and that the congruency of affective sound and affective meaning benefit from additional processing network. The corresponding neural mechanism potentially responsible for this sound-meaning interaction could be revealed in a brain area known for its role in supramodal emotion integration; i.e., the left amygdala.

By moving beyond the ideophone words as the only well-studied form of iconic mappings, our finding emphasizes that iconicity is not a marginal phenomenon in language. Some previous proposals restricted the role of iconicity to an earlier evolutionary stage of human language before the jump towards using a symbolic and arbitrary system, hence considering iconicity a living 'fossil' of proto-language. By showing the effect of iconicity in language processing and its underlying neural substrates, we, however, provided evidence for the presence, and for the neuropsychological reality of this phenomenon in today's language. These results point to the indispensable role of iconicity in building a comprehensive knowledge about human language, and encourage future research to incorporate the underestimated iconic constituent of verbal symbols into linguistic theories, and to revisit the predominantly arbitrary character of language "awaiting due consideration in modern linguistic methodology" (Jakobson, 1965).

9.6 Materials and Methods

9.6.1 Stimuli

120 nouns (one to three syllables long) were selected for a 2x2 design (30 words for each condition) characterized by an orthogonal twofold manipulation of *lexical* and *sublexical arousal*. For *lexical arousal* we used ratings of words' affective meaning (min = 1: very low arousing, max = 5 very high arousing) from the normative database BAWL-R (Võ et al., 2009). *Sublexical arousal* was calculated based on features extracted from the acoustic representation of words applying the psychoacoustic model developed in a previous work (Aryani, Conrad, et al., 2018). For this, words were uttered in a list-like manner by a professional male actor who was a native speaker of German and recorded with a sampling frequency of 48 kHz and 16 bits per sample. Audio files were then normalized to have the same loudness by matching their root-mean-square (RMS) power. Words were divided into two distinctive conditions of "high" and "low" arousing for each of the factors *lexical arousal* ('High' > 3.25, 'Low' <2.75) and *sublexical arousal* ('High' > 3, 'Low' < 3), and carefully controlled for relevant psycholinguistic variables across all of four cells of experimental conditions. *Lexical arousal* (and *lexical valence*) was closely controlled for between the two cells of *sublexical arousal*, and vice versa (Table1). In order to create an acoustic baseline, we randomly selected 16 words from the word material (4 from each condition) and converted them to signal-correlated noise (SCN; Schroeder, 1968). Along with our stimulus material (120 words + 16 SCN), a total of 76 additional words (mostly emotionally neutral) were presented which were a part of another study, and were discarded from further analysis here.

Variable	Word Category								Inferential Statistics
	HH		HL		LH		LL		
	M	SD	M	SD	M	SD	M	SD	
Lexical Arousal	4.07	0.24	4.04	0.22	1.99	0.16	1.99	0.18	F(3,116)=983, p<0.0001
Lexical Valence	-1.83	0.52	-1.83	0.51	0.22	0.36	0.18	0.37	F(3,116)=205, p<0.0001
Sublexical Arousal	3.36	0.31	2.76	0.19	3.30	0.27	2.77	0.21	F(3,116)=50.5, p<0.0001
Word Frequency	0.64	0.75	0.74	0.76	0.57	0.78	0.51	0.75	F(3,116)=0.47, p=0.69
Imageability Rating	4.78	1.01	4.56	1.0	4.93	0.90	5.02	1.16	F(3,116)=1.11, p=0.34
# Syllables	1.86	0.73	2.1	0.54	2.0	0.69	2.03	0.61	F(3,116)=0.68, p=0.56
# Phonemes	5.3	1.36	5.23	1.10	5.13	0.89	4.93	1.20	F(3,116)=0.57, p=0.63
duration (ms)	873	116	850	102	826	108	836	100	F(3,116)=1.06, p=0.36

Table 9.1. Characteristics of word stimuli. HH= High-High, HL=High-Low, LH=Low-High, LL=Low-Low: the first letter indicates the *lexical* and the second *sublexical arousal*.

9.6.2 Participants

Twenty-nine right-handed German native speakers (17 women, mean age 25.2 years, range: 20-35 years) with no history of neurological or psychiatric illness or any hearing problems volunteered to participate in the study, receiving either 15 Euros or psychology course credit for their participation. The Ethical Committee of the Freie Universität Berlin had approved the investigation. In2formed consent was obtained according to the Declaration of Helsinki.

9.6.3 Procedure

Spoken words were presented via MRI-compatible headphones sufficiently shielded from scanner noise to ensure clear perceptibility. Participants were instructed to pay attention and to carefully listen to the words. A trial began with the presentation of a fixation cross for between 1500 ms and 6500 ms, jittered in steps of 500 ms, in the center of the screen. Jittering durations and the stimulus presentation order over different experimental conditions (HH, HL, LH, LL, SCN, Fillers), were optimized with Optseq2 to ensure a maximal signal-to-noise ratio (Greve, 2002). After presentation of a stimulus the fixation cross disappeared. All blocks were set to a fixed length of 370 volumes. A total number of 10 trial words were presented prior to the experiment, which were excluded from the analysis. Words were split and presented in two runs. Between the two runs the participants could take a break.

9.6.4 fMRI data acquisition

Imaging data were collected on a Siemens Tim Trio 3T MR scanner. Functional data used a T^{*}_{2}-weighted echo-planar sequence [slice thickness: 3 mm, no gap, 37 slices, repetition time (TR): 2 s, echo time (TE): 30 ms, flip angle: 70°, matrix: 64 × 64, field of view (FOV): 192 mm, voxel size: 3.0 mm × 3.0 mm × 3.0 mm, 2 x 305 volumes, acquisition time: 2x 610 s]. At the beginning of the experimental session, magnitude and phase images for the field map were acquired: [slice thickness: 3 mm, no gap, 37 slices, TR: 488 ms, 2 TE: 4.92 and 7.38 ms, flip angle: 60°, matrix: 64 × 64, FOV: 192 mm, voxel size: 3.0 mm × 3.0 mm × 3.0 mm, acquisition time: 65 s]. Individual high-resolution T1-weighted anatomical data (MPRAGE sequence) were also acquired (TR: 1.9, TE: 2.52, FOV: 256, matrix: 256 × 256, sagittal plane, slice thickness: 1 mm, 176 slices, resolution: 1.0 mm × 1.0 mm × 1.0 mm).

9.6.5 Post-scan tests

Unannounced recognition test: Participants were presented with the same 120 words used in the scanner (OLD) mixed with 120 new words (NEW) which were matched with OLD items for word frequency, number of letters, number of phonemes, number of syllables, and imageability rating, as well as valence and arousal (selected from the same range as used for OLD items). Participants were asked to rate how confident they were that the presented word was or was not part of the word list in the scanner (from certainly not presented in the scanner = 1 to certainly presented in the scanner = 5).

Ratings: For rating of *lexical* and *sublexical arousal*, the same instruction was used as in the original rating study of the BAWL (Võ et al., 2009). For rating of *sublexical arousal*, participants were additionally instructed to only concentrate on the sound aspect of the words while trying to suppress their meaning (cf. Aryani, Conrad, et al., 2018).

9.6.6 fMRI Preprocessing

The fMRI data were preprocessed and analyzed using the software package SPM12 (www.fil.ion.ucl.ac.uk/spm). Preprocessing consisted of slice-timing correction, realignment for motion correction, magnetic field inhomogeneity correction through the creation of a field map, and coregistration of the structural image onto the mean functional image. The structural image was segmented into gray matter, white matter, cerebrospinal fluid, bone, soft tissue, and air/background (Ashburner & Friston, 2005). A group anatomical template was created with DARTEL (Diffeomorphic Anatomical Registration using Exponentiated Lie algebra; Ashburner, 2007) toolbox from the segmented gray and white matter images. Transformation parameters for structural images were then applied to functional images to normalize them to the brain template of the Montreal Neurological Institute (MNI) supplied with SPM. Functional images were spatially smoothed with a kernel of 6 mm full-width-at-half-maximum during normalization.

9.6.7 fMRI Analysis

GLM Analysis

Voxel-wise fixed effects contrast images made by subtraction analyses were performed at the single subject level and random effects analyses (Holmes & Friston, 1998) were conducted at the group level to create SPM contrast maps. On the single-subject level, each of the six conditions (HH, HL, LH, LL, SCN, and FILLERS) was convolved with the haemodynamic response function (HRF). Events were modeled as delta functions with zero duration. The beta images of each conditional regressor were then taken to the group level, where a full-factorial 2nd level analysis with the factors *lexical arousal* and *sublexical arousal* was used. An unconstrained non-directional 2x2 ANOVA whole brain analysis was performed with the factors *lexical arousal* (High, Low) and *sublexical arousal* (High, Low), to investigate the overall presence of main and interaction effects. For whole-brain fMRI analyses, we used the cluster defining threshold (CDT) of $p < 0.005$, then applied cluster-level family-wise error (FWE) correction to $p < 0.05$ for the entire image volume, as suggested by Liebermann and Cunningham (2009) for studies in cognitive, social and affective neuroscience. As recent evidence suggests that this may be too liberal (Ekund et al., 2016), in order to reduce the likelihood of false positives and to increase the specificity of the results, we repeated the analysis with CDT < 0.001 (FWE-corrected p-values < 0.05) and reported both results.

The labels reported were taken from the 'TD Labels' (Lancaster et al., 2000) or 'aal' labels in the WFU Pickatlas Tool. The Brodmann areas (BA) were further checked with the Talairach Client using nearest gray matter search after coordinate transformation with the WFU Pickatlas Tool.

PPI Analysis

To investigate regions showing significant functional connectivity with brain regions processing sound and meaning of words related to iconicity, generalized psychophysiological interactions (gPPI) (McLaren, Ries, Xu, & Johnson, 2012) were analyzed between the activations in seed regions representative for acoustic and semantic processing. For the sound aspect, we selected the left superior temporal gyrus (STG), and in particular auditory cortex, as a representative seed region for acoustic processing. For the meaning aspect, we focused on the left inferior frontal gyrus (IFG) due to its clear involvement in semantic processing (Binder, Desai, Graves, & Conant, 2009) and in appraisal of the words' affective meaning (Kuhlmann et al., 2016). These seed regions were defined based on the observed activations in the comparison of all words and acoustic baseline (Words > SCN). The left STG was extracted from a cluster of activation with an activation peak in [x y z] = [-59 -14 3], and a cluster size of k = 437 voxels, and the left IFG from a cluster of activation with an activation peak in [x y z] = [-41 33 -14], and a cluster size of k = 432 voxels (see

Results). To analyze patterns of functional connectivity, we used the gPPI toolbox which produces a design matrix with three columns of condition-related onsets with canonical HRF, BOLD signals deconvolved from the seed region, and PPI regressors at the individual level. Thus, the GLM of this analysis included PPI and condition regressors of the four experimental conditions (i.e., HH, HL, LH, LL).

Chapter 10

Extracting salient sublexical units from written texts: EMOPHON, a corpus-based approach to phonological iconicity [6]

10.1 Abstract

A growing body of literature in psychology, linguistics, and the neurosciences has paid increasing attention to the understanding of the relationships between phonological representations of words and their meaning: a phenomenon also known as phonological iconicity. In this article, we investigate how a text's intended emotional meaning, particularly in literature and poetry, may be reflected at the level of sublexical phonological salience and the use of foregrounded elements. To extract such elements from a given text, we developed a probabilistic model to predict the exceeding of a confidence interval for specific sublexical units concerning their frequency of occurrence within a given text contrasted with a reference linguistic corpus for the German language. Implementing this model in a computational application, we provide a text analysis tool which automatically delivers information about sublexical phonological salience allowing researchers, inter alia, to investigate effects of the sublexical emotional tone of texts based on current findings on phonological iconicity.

[6] This chapter is submitted as: Aryani, A., Conrad, M., & Jacobs, A. M. (2013). Extracting salient sublexical units from written texts: "Emophon," a corpus-based approach to phonological iconicity. *Frontiers in Psychology*, *4*, 654.

10.2 Introduction

How do literary texts affect the reader? Can sounds have intrinsic, autonomous meaning, particularly in literary and poetic language?

The Russian Formalists were the first to examine these questions in a systematic fashion. They took the position that the phonological structure of poetry has a function beyond the decorative, and should be an object of study in its own right. In their examination of sound in poetry, the Formalists went far towards answering the questions of how sound patterns are organized in verse and what onomatopoetic types of sound patterns may be identified. Their approach provided a procedural basis for subsequent investigations of sound in poetry using objective criteria and a linguistic focus (Jakobson, 1960; Shklovsky, 1925/1990; Trotsky, 1957; see Mandelker, 1983 for a review).

Over the last two decades, the study of the aestheticized form of language (e.g., literature, poetry) has been complemented by developments in cognitive linguistics, and particularly by research into an area at the interface between linguistics, literary studies, and cognitive psychology known as 'cognitive poetics' (e.g., Gavins & Steen, 2003; Miall & Kuiken, 1994; Stockwell, 2002; Tsur, 1992a), 'cognitive stylistics' (e.g., Semino & Culpeper, 2002; Semino, 2009), and 'neurocognitive poetics' (e.g., Jacobs, 2011; 2013; 2014). This interdisciplinary field offers (neuro-)cognitive hypotheses to relate "the specific effects of poetry" to "the particular regularities that occur in literary texts" in a systematic way (Tsur, 2003). In contrast to other areas of literary studies, (neuro-)cognitive poetics deals with the central question of how readers comprehend and interpret the language of literary texts by conducting experiments "with different types of literary discourse, in different reading contexts with different kinds of readers" (Van Dijk, 1979).

In this article, we call upon the theoretical framework of (neuro-)cognitive poetics, particularly those insights emphasizing the figurative and embodied nature of linguistic knowledge, in order to investigate the relationship between the linguistic form of a text and the part of its meaning or emotional content, termed 'iconicity'. We discuss the feasibility of iconicity being a potential indicator of the emotive qualities of a literary text specifically as evoked by particular phonemic structures. The basic idea behind such emotional impact of a literary text lies in the perceived similarity between the semantic content of the text and iconic associations based on the phonological salience at the level of the whole text as assumed by foregrounding theory (e.g., Garvin, 1964; van Peer, 1986; Miall & Kuiken, 1994; Hakemulder, 2004).

Our measure of phonological salience is based on the deviation of the observed frequency from the expected frequency of particular phonemes or phonemic clusters in a given text. It's worth noting that the term 'salience' used in this study refers to a technical term from literary analysis, and indicates an over-proportionally use of specific elements in a given text. We do not claim that our definition of the term 'salience' always provides a perfect match to psychological salience in terms of how readers' perception of literary text might be affected by salient items. But still, our statistical operationalization might provide a good proxy for it in most cases. The focus of this study is on reliably extracting such phonologically salient units from written texts as a quantitative method facilitating further investigations in (neuro-)cognitive poetics.

10.2.1 Foregrounding as a stylistic strategy

Foregrounding refers to a form of textual patterning and is capable of working at any level of language such as phonology, morphology, syntax, and semantics. It is essentially used as a stylistic technique to - from Shklovsky's Russian term *ostranenie* - defamiliarize the reading experience in textual composition through stylistic distortion of some sort, either through an aspect of the text which deviates from a linguistic norm or, alternatively, where an aspect of the text is brought to the fore through repetition or parallelism (Simpson, 2004). The analogy or correlation between the Gestalt notion of figure-ground (Rubin, 1921; Köhler, 1929; Koffka, 1935) and the idea of foregrounding theory in literature has led to applying the Gestalt principles to literary studies. The concept of figure-ground perception in Gestalt psychology is a fundamental organizing principle assuming that the perceiver distinguishes between incoming visual sensations and notices some as 'more salient figures' in front of a 'less salient ground' (Wallace, 1982). Figures, in literary texts, could be observed and classified in the sense of deviant or foregrounded features of literary language (Leech, 2008). However, only after empirical results have shown that literary foregrounded elements are related to a more intensive and extensive cognitive processing (van Peer, 1986), increased memory for stylistic features (Zwaan, 1993) and deeper emotional experience (Miall & Kuiken, 1994; Hakemulder, 2004), the concept of foregrounding in modern text processing psychology has been seen "as a manifestation of the Gestalt psychological principle of figure-ground discrimination" (van Holt & Groeben, 2005). Besides empirical evidence from behavioral studies, recent functional neuroimaging experiments have provided supportive evidence for foregrounding theory and have shown that foregrounded items in a text can set the reader into a mode of aesthetic perception (Bohrn et al., 2012; 2013). Comparing familiar with defamiliarized proverbs, Bohrn and colleagues have shown that defamiliarized proverbs increase cognitive processing demand and activate brain areas related to affective processing (orbitofrontal cortex, amygdala).

At the phonological level, foregrounded elements such as alliteration or rhyme can cause a "conscious sub-vocalization" and, at the same time, "aesthetic feelings, interest, curiosity, pleasure and self-reflection" (Jacobs, 2011). By means of four studies, Miall and Kuiken (1994) collected segment by segment reading times and ratings from readers of three different short stories which contained a variety of foregrounded features (i.e., phonological, grammatical and semantic). They showed that the degree to which foregrounding is present in the segments of a story is a predictor of both reading times and readers' judgments of 'strikingness' and 'affect' and, interestingly, that the effective foregrounded elements were primarily phonological and semantic.

This assumed enhancing aesthetic and emotional effect of foregrounding on reading experience, combined with the semiotic notion of iconicity, which is discussed in the next section, will provide a theoretical framework to allow the computation of (sublexical) emotive qualities of a literary text by means of phonological foregrounded elements and their iconic associations.

10.2.2 Iconicity in language and literature

In his semiotic theory of the sign, Charles S. Peirce (1931) argued that the sign (representamen) can refer to its object through relationships of *similarity, contextual contiguity*, or *law*. According to this trichotomy, the sign is called an *icon*, an *index* or a *symbol* respectively. An Icon, or what is also called 'a motivated sign' (Nöth, 2001), exhibits some relevant properties of the object; its form is similar to, and shares qualities with its referent. According to Peirce, there are three categories of iconicity: imaginal, diagrammatic, and metaphoric iconicity, with an increasing degree of abstraction related to their representing object. A well-known form of imaginal iconicity is the use of onomatopoeia, in which the referent is an acoustic signal and the link between sign and object is direct. In metaphoric iconicity, however, this link is indirect and associative (De Cuypere, 2008), i.e., a sensation, a feeling (e.g., happiness) or a property (e.g., size, color). The psychological aspect of the process of language comprehension, in the iconic case, "can be conceived of as the vicarious experiencing of events in the real world" (Segal, 1995; Zwaan et al., 2001; Zwaan & Yaxley 2003).

By applying Peirce's iconicity to literary studies and poetry, Jakobson (1966) added instances of iconicity in syntax, phonology and morphology, and exemplified the similarity between form and meaning by the famous Caesar's phrase, "veni, vidi, vici" which mirrors "the order of narrated events in time or in rank" by means of syntactical and phonological icons such as cadence and repeating vowel sounds of the words. In his model of the functions of language, Jakobson distinguishes six elements or factors that are necessary for communication to occur. In poetry, according to this model, the emphasis is on the form of the message itself , and language not only serves as a medium of expression but is also "set toward the message as such, focus on the message for its own sake" (Jakobson, 1960). That is, the *differentia specifica* of poetry is located in its formal characteristics and iconic properties.

In the definition of the arts by Susanne K. Langer (1953) as 'the semblance of felt life', the relation of 'form' to 'feeling' lies at the basis of poetry and all art. This view has been expanded by Margaret H. Freeman (2009) who suggests that 'form', ‘feeling’ and even ‘meaning’ are all intertwined components of the cognitive processes of the embodied human mind in language and literature. Iconicity is an instructive example for the integration of these components in language - it creates “sensations, emotions and images that enable the mind to encounter them as phenomenally real” (Freeman, 2009). Lakoff and Turner, consistently, commented on the role of iconicity as “metaphorical image-mapping in which the structure of the meaning is understood in terms of the structure of the form of the language presenting the meaning” (Lakoff & Turner, 1989). Such image-mapping, according to them, is enabled by image-schemas which are formed from our embodied interactions. This embodied perspective of nature of icons, as a part of linguistic signs, challenges the simplified notion of the arbitrariness of linguistic signs.

10.2.3 Linguistic signs: embodied or symbolic?

As language has been commonly considered symbol system in which meaning is constructed through formal symbol manipulation, the following question arises: how can purely arbitrary symbols operate with non-arbitrary embodied signs (e.g., icons)? The poetic use of language might be the best example of the convergence of two seemingly contrasting approaches to language comprehension; one a purely ‘symbolic’ approach emphasizing the computational nature of symbols, and the other an ‘embodied’ approach postulating the grounding of symbols in the physical world (Barsalou, 1999; Glenberg, 1997; Pulvermüller, 1999). But for all that, is language comprehension either solely symbolic or solely embodied?

Based on Peirce’s trichotomy (icons, indices, symbols), Deacon (1997) suggests a hierarchical relationship of signs and argues that each sign is built from combinations of signs represented at its lower level (e.g., indices can be built from icons, symbols from indices). He claims that signs have the aptitude of operating not only with the signs in their own level but with those from other levels. Deacon proposes that the evolutionary jump to using a symbolic system from an iconic and indexical system and the ability to make links between symbols as well as between symbols and indices and icons can explain why humans have language (Deacon, 1997). This idea of a hierarchical relationship between signs is part of ‘the symbol interdependency hypothesis' which attempts to reach an agreement between two apparently contrasting accounts of language comprehension: symbolic and embodied (Louwerse, 2007; 2011; Louwerse & Jeuniaux, 2008). According to the symbol interdependency hypothesis, language comprehension is both symbolic and embodied. It can be symbolic by bootstrapping meaning through relations between the symbols, but it can also be embodied through the dependencies of symbols on indices and icons. This means that language comprehension is not solely symbolic because comprehenders can always activate embodied representations (e.g., indices and icons). The basic idea of the symbol interdependency hypothesis can be tracked back to Jakobson's view of the interrelatedness of sound and meaning. Jakobson supported a similar position to iconic and indexical use of language and proposed that "the iconic and indexical constituents of verbal symbols have too often remained underestimated or even disregarded; on the other hand, the predominantly symbolic character of language and its subsequent cardinal difference from the other, chiefly indexical or iconic, sets of signs likewise await due consideration in modern linguistic methodology". (Jakobson, 1966)

In the next section we will briefly outline some supportive results from studies in experimental psychology and linguistics on iconicity to account for its feasibility for extracting the emotional tone of a given text. Our focus will be on the relationship between the sound of words (phonological representations) and their meaning or emotional content; a phenomenon known as phonological iconicity that challenges the principle of arbitrariness - one of the tenets of Ferdinand de Saussure's theory of linguistic signs (Saussure, 1916).

10.2.4 Studies in phonological iconicity

Besides onomatopoetic words, which sound like the concept they describe (e.g., click, bang, splash, boom), phonaesthemes are an instructive example for the relationship between sound and meaning (Schrott & Jacobs, 2011). Phonaesthemes are sounds, sound clusters, or sound types that are directly associated with a lexical category or meaning. The initial cluster /gl/ is often cited as an example of an English phonestheme. It occurs in many words used for 'shiny things': glisten, gleam, glint, glare, glam, glimmer, glaze, glass, glitz, gloss, glory, glow, and glitter (Wallis, 1699; Magnus, 2001, Bergen, 2004). Likewise, in German, nouns starting with /kno/ and /knö/ are mostly small and round: Knoblauch 'garlic', Knöchel 'ankle', Knödel 'dumpling', Knolle 'bulb', Knopf 'button', Knorren 'knot (in a tree)', Knospe 'bud', Knoten 'knot (in string or rope)' ("Sound Symbolism," n.d.).

The idea of phonological iconicity has ancient roots. In Cratylus (Plato, 1892), Plato has Socrates propose that the foundation of word semantics must lie in phonology and the way they sound. In contrast to de Saussure and Hjelmslev, many influential linguists supported the position of possible synchronic and productive effects of a word's sound on its meaning. For instance, Jakobson proposed that, "the intimacy of connection between the sounds and the meaning of a word gives rise to the desire of speakers to add an internal relation to the external relation, resemblance to contiguity, to complement the signified by a rudimentary image" (Jakobson, 1937). That is, the effect of sound in the mind completes its meaning and this, according to Jespersen, may lead to a kind of "natural selection" that "makes some words more fit to survive" (Jespersen, 1922).

Since the 1920s, there has been a considerable amount of increasingly sophisticated experiments to test the functioning of sound symbolism in languages. Sapir (1929), for example, raised the issue whether phonemes in isolation are symbolic of differing size by using two nonsense words MAL and MIL. Subjects consistently judged MIL to denote a small object and MAL to a large object. Building on this study, Newman (1933) further investigated the symbolic connotations of nonsense words differing (by pairs) only in one vowel and asked the subjects, which of the pair seemed larger, smaller, darker, or lighter. He concluded that tongue position in articulation is basic in the patterning of both magnitude and brightness categories.

More recently, a growing amount of literature has paid significant attention to understanding the role of phonological iconicity in language. Nygaard and colleagues (Nygaard, Cook, and Namy, 2009) have reported that when native English speakers were presented with unfamiliar Japanese words, their ability to link these words to

English meanings was biased due to sound-meaning relationships of Japanese words (see also Nygaard, Herold and Namy, 2009). In a series of computational investigations, artificial language learning studies, and corpus analyses of English and French, Monaghan, Christiansen, and Fitneva have investigated arbitrary and systematic mappings between word forms and word meanings with regard to their respective advantages of word learning (Monaghan, Christiansen, & Fitneva, 2011). They have shown that systematicity facilitates learning to group words into categories whereas arbitrariness facilitates learning specific word meanings (see also Christiansen & Chater, 2008). Several studies from experimental psychology and linguistics have suggested the existence of a systematic relationship between sound and meaning (Bergen, 2004; Westbury, 2005; Otis and Sagi, 2008; Berlin, 1994; Cassidy, Kelly, & Sharoni, 1999, see Perniss, Thompson, & Vigliocco, 2010 for a review). A considerable demonstration for a non-complete arbitrariness between sound and meaning is that both adults and infants map nonsense words with rounded vowels (e.g., bouba) to rounded shapes and nonsense words with unrounded vowels (e.g., kiki) to angular shapes (Köhler, 1947; Ramachandran and Hubbard, 2001; Maurer, Pathman et al., 2006; Ozturk, 2013).

Apart from this general sound-meaning relationship, studies have explored a relationship between the way words sound and the emotional content expressed by them. Fóngay (1961), for example, applied statistical methods to examine the different tone qualities in six aggressive and six tender poems by the Hungarian poet Petöfi. While /t/, /k/, and /r/ were more frequent in aggressive poems, /l/, /m/, and /n/ were more frequent in tender poems. Based on information theory, Bailey (1971) analyzed a group of texts (poems vs. prose) to show whether in certain literary texts the frequency of occurrence of some phonemes is higher than this frequency in prose text, and revealed some meaningful patterns such as a preference for voiced over voiceless consonants and a preference for back over front vowels related to the aesthetic quality of poems. Recent machine learning experiments have shown that words expressing the same emotion (e.g., happiness) have significantly more in common with each other than with words expressing other emotions (e.g., sadness), suggesting that happy words might indeed sound happy (Nastase et al., 2007). In a similar study on affective words in Romanian and Russian, Sokolova and Bobicev (2009) captured the word form similarity by classification of words according to their emotion tags. The results suggest that the form of words allows for a reliable classification of emotion.

Evidence for the psychological reality of this sound-meaning correspondence comes from behavioral experiments. Wiseman and van Peer (2003), for instance, revealed that when German and Brazilian participants were asked to produce fantasy words corresponding either to the emotions experienced at a wedding or at a funeral, they tended to use similar consonants for respective emotional states, and this was independent of their native language. While nasal sounds (/m/, /n/) were more frequently used for the expression of sadness (funeral), plosive sounds (/p/, /b/, /d/, and /t/) were better suited to the expression of happy feelings (wedding) (Wiseman & van Peer, 2003; see Auracher et al., 2011). In an attempt to examine the universality of this phenomenon, Auracher et al. (2011) asked German, Chinese, Russian and Ukrainian native speakers to assess the emotional tone of poems with extreme values (highest vs. lowest) concerning the ratio of plosive versus nasal sounds. Their results were consistent with the results of Wiseman & van Peer (2003) and confirmed the relationship between the emotional tone (happy vs. sad) of poems and the manner of articulation (plosive vs. nasal) for all examined languages. More recently, Myers-Schulz et al. (2013) have shown that the perceived emotional valence of certain phoneme combinations depends on the dynamic shift within the phonemes' first two frequency components, suggesting that certain strings of phonemes have a non-arbitrary emotional quality (e.g., /sa:/ is perceived as positive, and /za:/ as negative). In several analyses of English poetry and lyrics, Whissell has proposed that most of the basic sounds of English have emotional connotations attached to them (see, for instance, Whissell, 2003). She used the '*Dictionary of Affect in Language*' (Whissell, 1989) to validate these connotations and reported that, for example, the /l/ sound has positive and gentle connotations while /r/ and /g/ sounds have harsher ones.

10.2.5 Extracting the salient phonological units

In sum, the use of stylistic elements in a text, such as foregrounded phonological salience can enhance aesthetic and emotional effects on reading experience. On the other hand, the perceived similarity or analogy between the phonological forms of those salient elements and their meaning, or their emotional impacts on the reader, as assumed by the semiotic notion of phonological iconicity, can be a reliable source to predict a part of the emotional and aesthetic qualities of a given text. Now, what exactly makes a phoneme or any other sublexical unit within a given text, regardless of its genre, "salient"?

All the above mentioned theoretical claims or empirical approaches to phonological iconicity encounter the following problem: They assume that single sublexical units already relate to semantics below the lexical level in that they either possess an iconic value on their own, or directly convey meaning via a basic grounding of semantics or emotion at the level of single sounds. But unlike words, all these more basic elements supposed to carry such basic semantic or emotional 'meaning' almost necessarily have to form part of any complex speech signal – given the general limitations of the phoneme inventory of natural languages. Thus, how could a single sublexical unit serve as a linguistic sign conveying specific semantic meaning if it almost necessarily will be present in any text, regardless of its meaning? A plausible tentative answer to this question is: frequency.

Single sublexical units only gain such 'semantic sign' character if occurring with elevated frequency within a given text. Notably, this is the point of view adopted by approaches trying to assign emotional values to sublexical units on a purely empirical base (e.g., Bailey, 1971; Whissell, 2003; 2009; Auracher et al., 2011) establishing relationships between lexico-semantic word meaning and sublexical organization. But what kind of information regarding the frequency of occurrence of single sublexical units could be extracted from single texts standing in the focus of interest?, as this would be the case for all kinds of scientific approaches focusing on already existing language samples as literary productions. Here, the analysis of frequencies of occurrence for single sublexical units – in order to label them as "salient" or not –

faces two issues: i) The more trivial one of how to count them? ii) The less trivial one of when to label a given frequency of occurrence as salient?

10.3 Material and Methods

To extract the salient sublexical units within a text, we developed a probabilistic model. The basic idea is to weight the frequencies of occurrences of a sublexical unit in a text by comparison to a linguistic corpus serving as a reference. Both single phonemes and sub-syllabic segments (e.g., onset, nucleus, and coda) have been considered as relevant indications of phonological iconicity. The higher the discrepancy between the relative occurrences of sublexical units within a given text on the one hand and within the reference corpus on the other, the more 'significant' the specific use of such units may appear.

According to foregrounding theory, the foregrounded pattern in a literary text or in poetry deviates from a norm, either through replication or through parallelism. In his neurocognitive model of literary reading, Jacobs (2011; 2013) assumes that not the deviation by itself, but the ratio between deviation and standard, between foregrounded and backgrounded elements, is the crucial factor for emotional and aesthetic experiences of literary reading. To proceed on the assumption that these norms originate from everyday spoken language, a corpus that is to be used as a reference should be representative of everyday language.

10.3.1 Reference Corpus

Recent studies have shown that word frequency calculated from corpora based on films and television subtitles can better account for reading performance than the traditional word frequency based on books and newspapers, since the language used in subtitles greatly approximates everyday language (interactions with objects and other people); evidence comes from: German (Brysbaert et. al, 2011), Spanish (Vega et al., 2011; see also Duchon et al., 2013), Greek (Dimitropoulou et al., 2010), Chinese (Cai et al., 2010), and Dutch (Keuleers et al., 2010). Brysbaert et al. have shown that frequency measures should best be based on a corpus of at least 20 million words (Brysbaert et. al, 2011). They reported that the best quality of frequency measures for German regarding their correlation with behavioral word processing data is observed for SUBTLEX-DE; a corpus of 25 Million German words consisting of movie and television subtitles. Accordingly, we decided in favor of SUBTLEX-DE as a reference corpus.

As the idea of phonological iconicity takes account of the phonological feature of language rather than orthography, each text to be analyzed should be converted into phonetic notation before analyzing it. For this purpose, we chose the German text-to-speech synthesis system MARY (Schröder et. al, 2003), which has a modular design that allows the user to access the grapheme-to-phoneme part separately. MARY uses the SAMPA phonetic alphabet for German (Wells, 1997) and outputs the phonemic transcription in an internal, low-level markup language called MARY-XML. To phonemize the text automatically, an extensive lexicon deals with known words and a letter-to-sound conversion algorithm with unknown words.

Since the selected corpus SUBTLEX-DE is only available in orthographic form, it was necessary to translate also the whole reference corpus into phonemic notation. On that account, we have split the corpus into small units, each including approximately 30000 characters[7]. Translating more than 6000 units through MARY, parsing resulted XMLs and then integrating them in one dataset, we have generated a complete phonemized corpus for the German language; a useful resource for future research in the area of psycholinguistics.

The following steps of analyses will focus on relative syllabic position of phonemes. Since syllable separation for each word is marked in phonemized SUBTLEX it was possible to extract all syllables of existing words. Based on internal structure of a syllable (Fig. 10.1) and a list of all 19 vowels in German (16 monophthongs and 3 diphthongs) each sub-syllabic unit has been segmented and its frequency of occurrence in the corpus has been counted.

[7] The splitting was needed because of the limitation of MARY's server performance and its permanent need for online-lookup in lexicon.

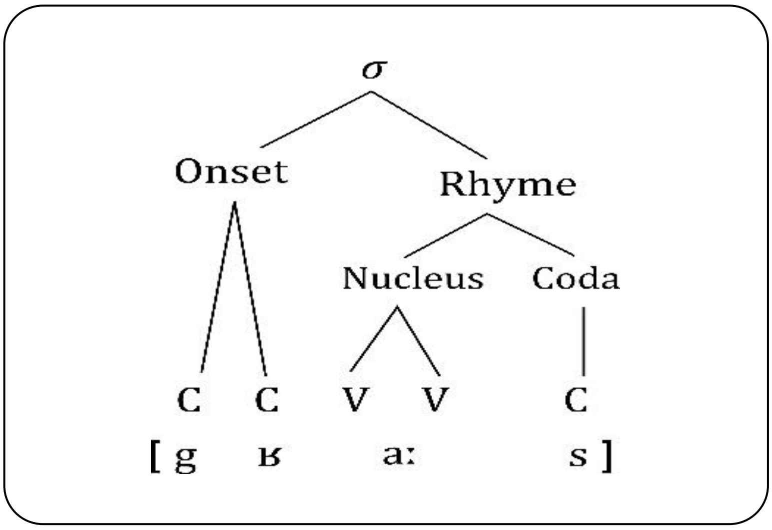


Figure 10.1. Representation of the structure of a syllable (σ) using the example of the German word 'Gras'.

Additionally, the sum of all sub-syllabic units and all phonemes existing in the corpus have been calculated in order to obtain the relative frequency of occurrence of each unit (Table 10.1).

phoneme	freq(%)	Onset	freq(%)	Nucleus	freq(%)	Coda	freq(%)
n	9.608	d	13.431	@	17.462	n	26.010
t	7.282	v	7.965	I	13.288	6	13.266
@	6.714	n	7.863	a	11.199	s	10.519
I	5.110	z	7.262	i:	8.667	C	7.894
6	4.761	t	6.792	E	7.637	t	4.930
s	4.535	m	6.792	aI	5.863	l	4.168
d	4.351	g	5.572	a:	5.467	st	3.793
a	4.306	b	5.358	e:	4.684	m	3.508
l	3.372	f	4.891	u:	4.023	nt	3.464
i:	3.332	h	4.756	U	3.947	Ct	2.378

Table 10.1. List of 10 most frequent phonemes and sub-syllabic units in SUBTLEX-DE (in SAMPA alphabet)

10.3.2 The probabilistic model

Since the significance of the use of a certain sub-syllabic unit depends on both its relative frequency of occurrence and the length of a given text, the intended model has to be able to predict the expected frequency and its standard deviation according to the corpus as a function of relative frequency and text length. Given the frequency of occurrence of a certain sub-syllabic unit in the corpus $f_{u,c}$, the text length n_{text} , and the corpus' length n_c , the expected frequency of occurrence of this certain unit in a given text would be:

$$E_{u,text} = f_{u,c} \cdot \frac{n_{text}}{n_c} \qquad (1)$$

For simplification, the term $f_{u,c}/n_{c,}$ which is the relative frequency occurrence of a certain sub-syllabic unit will henceforth be called as $f_{u,r}$. For calculating the standard deviation as a function of n_{text} and $f_{u,r}$, one approach consists of considering the occurrence of each sub-syllabic unit in each position of the text as a success/failure experiment with a certain probability. That is, we consider the whole text as a collection of gaps which could be filled by any sub-syllabic units contributing to the formation of the whole text. From this perspective, each given text could be seen as a sequence of $n_{text} \times$ yes/no experiments regarding each particular sub-syllabic unit; as if all units were competing for filling each gap with a specific probability (Fig 10.2).

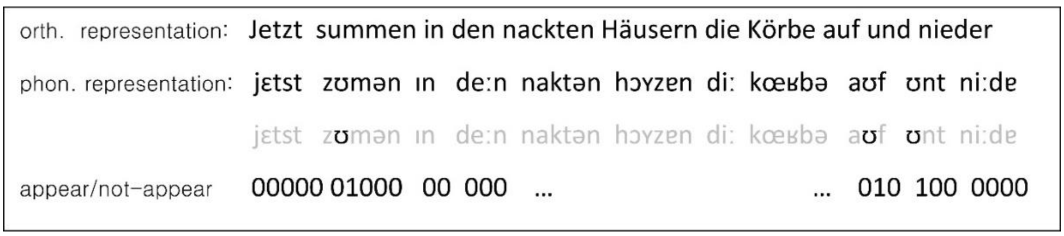


Figure 10.2. A random text is simulated as random binary sequences 000000100…. The symbol '1' appears with probability p and models a successful occurrence of a certain sub-syllabic unit in a text (in this example the phoneme ʊ), and the symbol '0' accounts for its unsuccessful occurrence with probability $1-p$.

Assuming that the occurrence of each unit in each position in the text is independent of the other units (which is not necessarily given for language) and that the probability that it succeeds in filling the position is equal to its relative frequency, one can consider the text as several Bernoulli trials with a binomial distribution. The standard deviation of a binomial distribution is:

$$u \sim B(n_{text}, f_{u,r}) \rightarrow \sigma u = \sqrt{n_{text} \times f_{u,r} \times (1 - f_{u,r})} \qquad (2)$$

Given the fact that only specific arrangements of sub-syllabic units can contribute to meaningful words, it is obvious that the distribution of sub-syllabic units in a text does not have completely random characteristics as preconditioned by a binomial distribution. However, this distribution can help to form a rough estimate about the characteristics of standard deviation and its dependency on both input factors (i.e., $n_{text}, f_{u,r}$).To calculate standard deviations more precisely, we chose an empirical approach by pulling numerous chunks of sub-syllabic units from the corpus. For a text with a certain length, a text sample with the same length is randomly pulled from the corpus and the frequency of occurrence of all sub-syllabic units is counted in this sample. Since the samples should be representative for a larger population, this procedure is repeated (with replacement) for 1 Million times for each specific length of text. It's worth pointing out that the corpus includes almost 25 Million words and more than 2 Million sentences. We opted to let all chunks start with the beginning of a sentence, because this best represents normal language. In consequence, 2 Million different chunks can be pulled from the corpus. For each set of pulling (i.e., 1 Million samples) the standard deviation for any sub-syllabic unit is calculated giving one point in the Cartesian coordinate which is representative for the value of standard deviation for a certain sub-syllabic unit dependent on the length of a given text (Fig. 10.3).

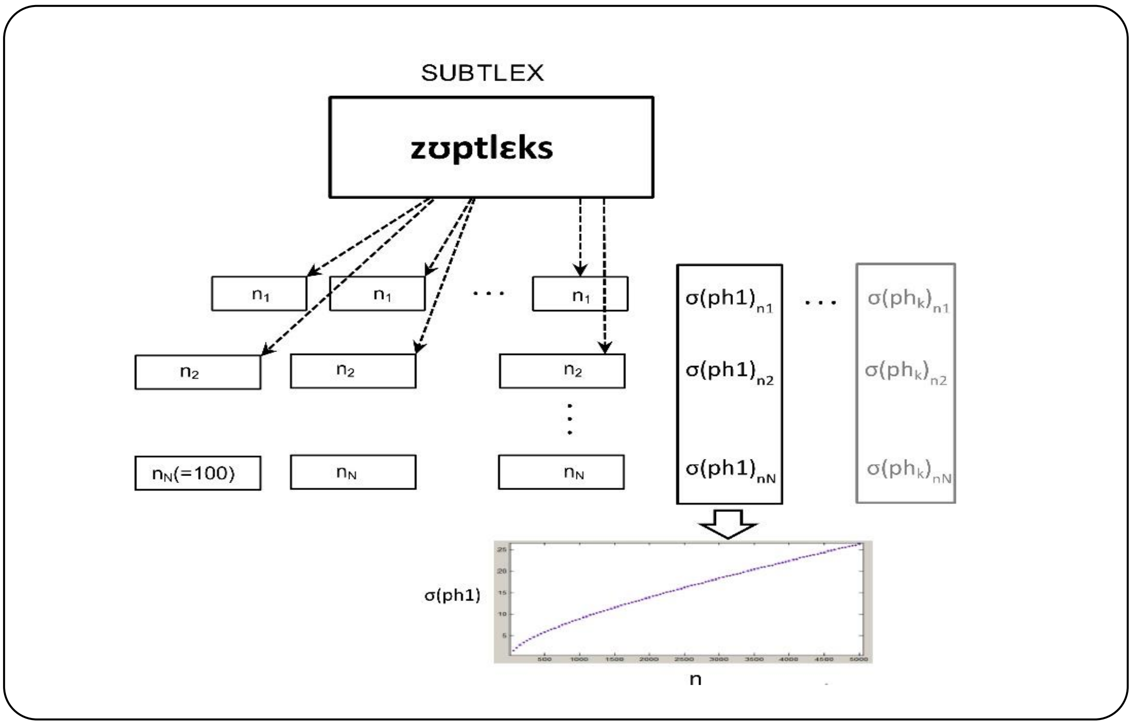
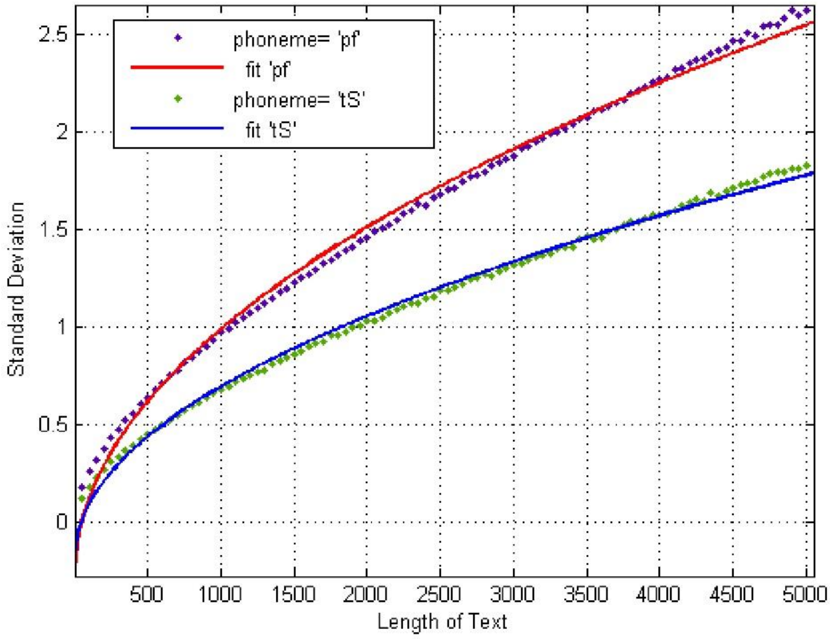


Figure 10.3. Calculation of standard deviation for each set of pulling. Note that the relative frequency of each sub-syllabic unit is constant giving one function for each unit.

This procedure of pulling is repeated for different lengths of texts, starting with the minimal size of 50 units for the first step, sampling 100 different lengths with steps of 50 units each (i.e., 50, 100, 150,…,5000) and making separate measurements on each series of texts with a certain length (Fig. 10.3). The analysis of all obtained curves shows, as expected, similar characteristics to a binomial distribution. Based on the mathematical equation for calculating the standard deviation of binomial distribution, we construct a curve that has the best fit to the series of our data points. In Figure 10.4 we represent standard deviation as a function of a text length using the example of the phonemes 'pf' (as in the German word 'Pferd') and 'tS' (as in the German word 'Deutsch').



Model:

fit(TextLength) = a*(sqrt(frequency*(1-frequency)*TextLength)) + c

Coefficients:

Phoneme	a	c
pf	1.672	-0.2738
tS	1.687	-0.1822

Model Goodness:

Phoneme	SSE	R-square	Adjusted R-square	RMSE
pf	0.226	0.9948	0.9947	0.04802
tS	0.09962	0.9952	0.9952	0.03188

Figure 10.4. Binomial-based model for prediction of standard deviation using the examples of ‘pf’ and ‘tS’.

Note that the curves in Figure 10.4 represent standard deviation as a function of text length alone, and not of relative frequency. In fact, each phoneme (as each sub-syllabic unit) has a certain relative frequency, which means, the factor frequency doesn’t appear as an input parameter, but rather as a constant value which is integrated in the function. In so doing, we could obtain a number of single functions each of which can predict the standard deviation of a certain sub-syllabic unit or a phoneme dependent on the text length. In order to obtain a general model for the prediction of any sub-syllabic units or phonemes, we compiled our all data in a three dimensional space as following: one dimension standing for the frequency of each sub-syllabic unit or phonemes (giving one point for each unit), one dimension standing for the text length (50, 100, 150,…, 5000), and one dimension for each standard deviation related to other two dimensions. In this three dimensional space and, again, based on binomial distribution, we constructed a surface for the best fit to our data (Figure 10.5).

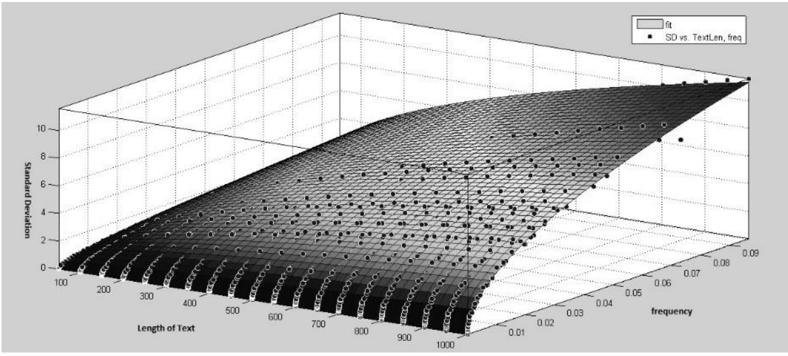


Figure 10.5. The 3D-model for prediction of standard deviation of sublexical units as a function of text length and frequency.

Though the accuracy of this latter model is less than accuracy of single functions for each unit, having a general model with text length and frequency as inputs simplifies the determination of standard deviation.

It's worth mentioning that this empirical approach is not needed for determining of expected value, because the calculation of expected value, which is based on formula 1, does not assume a particular type of distribution as it was the case for the calculation of standard deviation.

10.3.3 The text analysis tool "Emophon"

Based on the aforementioned model, we developed the computer linguistic tool Emophon which is a flexible tool for research and analysis of literary texts at the sublexical level. Emophon allows a step-by-step processing with an access to partial processing results. The core of the tool, as represented in Figure 10.6, consists of a segmenting module, the SUBTLEX-based model, and a comparing module. All sub-syllabic units and phonemes of a given text (phonemized through MARY) are found and segmented. By means of the frequency value of each unit (signed as f(Ph) in the representation) and the text length (calculated by tool), the integrated model provides a correspondent expected value and standard deviation for each unit. The number of each existing sub-syllabic unit and phoneme in the text is compared with the confidence interval provided by the model. Results are outputted as both graphic diagrams and numerical values of the degree to which the confidence interval is exceeded or not.

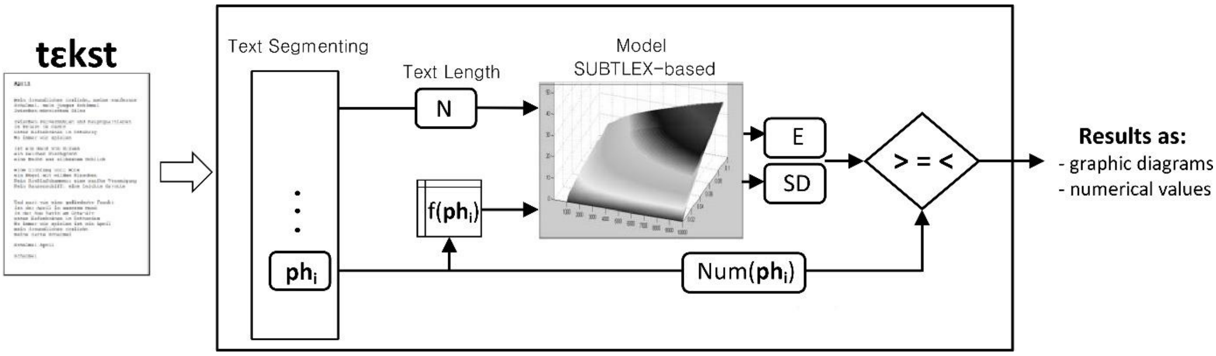


Figure 10.6. Representation of workflow and modules employed in the tool

A Graphic User Interface (GUI) allows user to browse and select texts. The sublexical salience in the text can be extracted and shown for structures at the following levels by simply selecting the corresponding option given in the GUI:

- whole phonemes in the text
- sub-syllabic onsets
- sub-syllabic nuclei
- sub-syllabic codas
- all single phonemes appearing in onsets
- all single phonemes appearing in codas

Emophon has been completely written in 'Python' (V2.5) and uses external libraries 'numpy' and 'mathplotpy'.

10.4 Results

In the following we will first demonstrate and validate the functionality of Emophon and then show how it can be used as a linguistic instrument for analyzing and extracting the phonological salient units in a text.

10.4.1 Validation of the functionality

To validate the functionality of Emophon we used a particular poem for which phonological deviation is known to some degree: the poem "Totenklage" written by German poet Hugo Ball (1886-1927); one of the leading poets of Dada movement and a pioneer of *sound poetry*. In sound poetry, phonological aspects of a poem are foregrounded. In this genre of poetry, linguistic meaning waives in whole or in substantial part and the language is purely formal and can be seen as mere sound material. The selected poem "Totenklage", in particular, is characterized by the frequent use of word-like units consisting of open syllables (CV; otherwise rather untypical for the German language) using preferentially specific vowels (see Figure 10.7).

Hugo Ball: Totenklage

ombula
take
bitdli
solunkola
tabla tokta tokta takabla
taka tak
Babula m balam
tak tru ü
wo um
biba bimbel
o kla o auwa
kla o auwa
la auma
o kla o ü
la o auma
klinga o e auwa
ome o-auwa
klinga inga M ao Auwa
omba dij omuff pomo auwa
tru-ü
tro u ü o a o ü
mo-auwa
gomun guma zangaga gago
blagaga
szagaglugi m ba o auma
szaga szago
szaga la m blama
bschigi bschigo
bschigi bschigi
bschiggo bschiggo
goggo goggo
ogoggo
a o auma

Figure 10.7. Hugo Ball's poem "Totenklage"

Since the “words” used in this poem are artificial, one might expect a high degree of phonological deviation from the German language and, in consequence, a large number of phonemes labeled as salient by the tool. In particular, we expect an exceeding of the confidence interval for the two vowels; /a:/ (as appeared in: tabla, tokta, taka, babula, biba, kla) and /o:/ (as appeared in: solunkola, wo, o, ome, omba, pomo) as well as for the consonant /g/ (as appeared in: gomun, guma, gago, goggo, bschiggo) because of their apparent high frequency of occurrence in the text. The poem was first converted to phonetic notation by using MARY. After reading of the phonemized text by the tool, one can optionally select one of the above-mentioned levels of seeking salience in the text (phonemes, onsets, nucleus, etc.). The graphical demonstration of results concerning the phoneme level is presented in Figure 10.8. The number of existing phonemes, the expected value and the confidence interval for each phoneme in the text (based on calculated standard deviation) are represented in the diagram. Phonemes that were significantly more frequent than expected and consequently exceeded the confidence interval are signed with an asterisk.

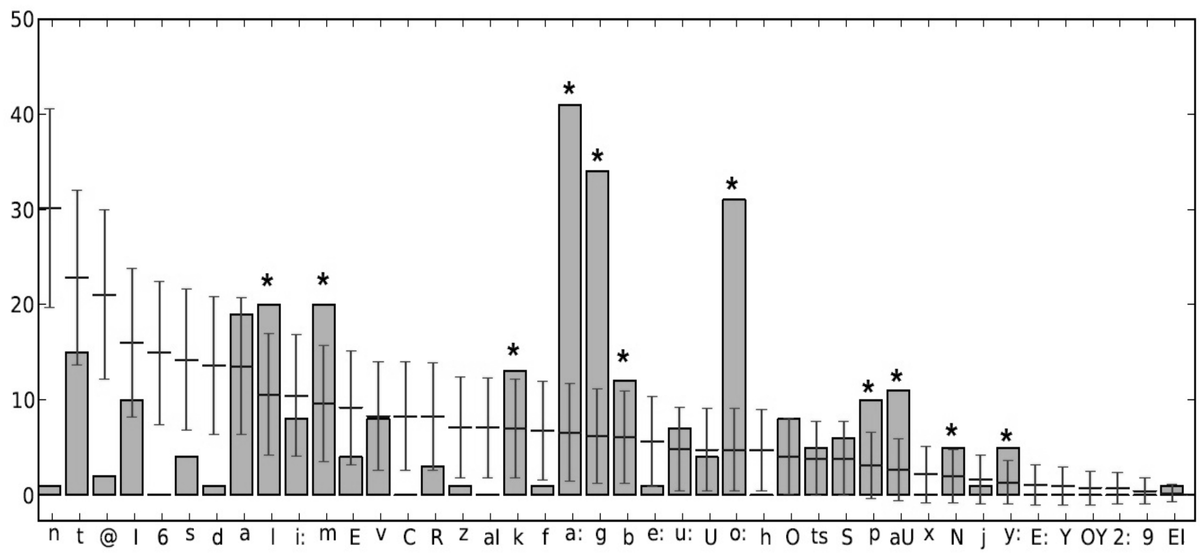


Figure 10.8. Diagram of phonemes in the poem “Totenklage” as outputted by the tool.

In addition, the exact degree to which each phoneme exceeds the confidence interval is outputted by the tool and represented in Table 10.2.

Table of Salient Phonemes

Phoneme	Tokens of this Phoneme	Tokens of Phonemes outside CI	UPPER /LOWER
a:	41	29.32	UPPER
g	34	22.87	UPPER
o:	31	21.94	UPPER
aU	11	5.13	UPPER
m	20	4.26	UPPER
p	10	3.39	UPPER
l	20	3.02	UPPER
y:	5	1.36	UPPER
b	12	1.05	UPPER
k	13	0.76	UPPER
N	5	0.17	UPPER
h	0	-0.40	LOWER
f	1	-0.62	LOWER
z	1	-0.84	LOWER
aI	0	-1.81	LOWER
C	0	-2.60	LOWER
s	4	-2.85	LOWER
d	1	-5.42	LOWER
6	0	-7.39	LOWER
@	2	-10.19	LOWER
n	1	-18.70	LOWER

Table 10.2. Numeric representation of the salient phonemes in the poem “Totenklage”. Note that fractions appear since the value of confidence interval is calculated based on the mathematical model.

As predicted, all of three mentioned phonemes (i.e., /a:/, /o:/, /g/) were marked as salient and positioned in the top of the salient list (Table 10.2, and Fig. 10.8). Notably, in addition, the phoneme /aU/ (as in the “words” ‘auwa’ and ‘auma’) largely exceeds its respective confidence interval (5.13 out of 11) though its absolute number of occurrence in the text is rather low. This is due to the fact that the expected value and the confidence interval for each phoneme positively correlate with its relative frequency in the corpus, which is low in this case. Similarly, by comparing the statistical salience of the phonemes /n/ and /d/ (with identical frequencies of occurrence =1), it becomes obvious that /n/, which has a higher relative frequency in the corpus and, accordingly, a higher expected value, is marked as more salient – due to its unexpected low actual occurrence in the text – than /d/ (18.7 for /n/ vs. 5.42 for /d/, less than expected). For the same reason, the phoneme /e:/, which has the same frequency of occurrence as /n/ and /d/ isn't marked by the tool as salient at all, by virtue of its very low relative frequency in corpus (Fig. 10.8).

Similar analyses can be conducted at the all above-mentioned levels to extract the sub-syllabic salient units in the text. Figure 10.9 demonstrates results of such an analysis for sub-syllabic onsets for the same poem. Similar to the previous analysis of phonemes, a large portion of existing onsets is marked as salient. In addition to the salient phoneme /g/ from the previous analysis, the onsets /pS/, /kl/, /bl/, and /tr/ show a large degree of salience in the poem (the exact numeric values of saliency are not shown due to the limited place).

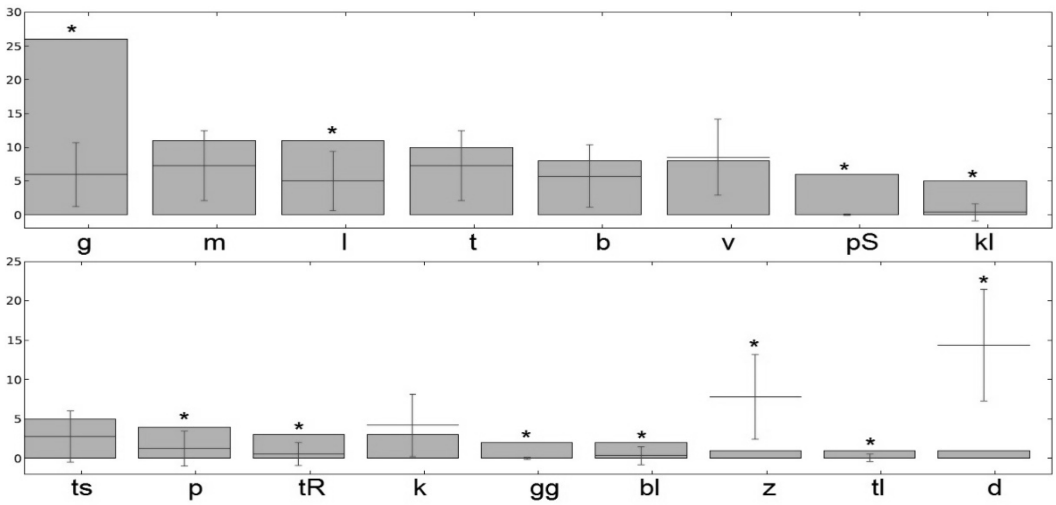


Figure 10.9. Diagram of "salient" syllabic onsets in the poem "Totenklage" as outputted by the tool.

10.4.2 Testing an exemplary hypothesis

To illustrate the potential use of Emophon as a linguistic instrument for analyzing the phonological salient units in texts following an exemplary hypothesis, we focus on the theory of foregrounding assuming that poetic language deviates from norms characterizing ordinary language use. We hypothesize that this deviation is observable at the phonological level and that it can be measured by the numbers of salient phonological units as provided by our tool. To this end, we chose 20 classical poems of approximately equal length (number of characters with space) from different German poets. All these poems are characterized by the two classical patterns of metric alignment and syllabic rhymes at the end of two lines either following an AABB or an ABAB pattern. In particular, these syllabic rhymes should lead to an increase in frequencies of occurrence of syllabic nuclei and codas in these examples of lyrical language when compared to everyday common language use.

As control texts, we chose 20 text passages - matching the lyrical texts in length - that had appeared in different online German newspapers as respective first articles on the day of analysis (see Appendix for a complete list of poems and newspaper articles). The subjects of these articles vary topically ranging from political to local news. After converting all texts in phonetic notion and analyzing them by the tool, we documented phonological salient units at 4 different sublexical levels: at the level of 1) phonemes, 2) sub-syllabic onsets 3) sub-syllabic nuclei, and 4) sub-syllabic codas. For a statistical comparison between the two groups of texts, we defined two measures indicating the degree of phonological salience for each text: 1) the number of phonological units being salient (i.e., the number of rows in Table 10.2 for the previous text), and 2) the absolute sum of all segments positioned outside the confidence interval (i.e., the absolute sum of the numbers in the 3rd column of Table 10.2). Thus, we obtained 8 indicators for each text, all of which can serve to compare poems and newspaper articles. The values of these indicators for each text according to the tool performance are represented in Table 10.3.

Applying t-tests on these 8 indicators, we assessed the likelihood that the means for the two types of texts (poems vs. newspaper) are sampled from the same sampling distribution of means. The results revealed a significant effect of text's type on the number of salient phonemes, t(38)=3.18, p=0.003, on the number of salient nuclei, t(38)=3.56, p<0.001, and the number of salient codas, t(38)=2.76, p=0.008, with each time more salient units being present in poems than in prose texts. Similarly to the number of salient units, there was a significant effect of text's type on the absolute sum of segments positioned outside the confidence interval: for phonemes, t(38)=2.2, p=0.033, for nuclei, t(38)=2.35, p=0.023, and for codas, t(38)=2.47, p=0.017, again, with more salient segments in poems than in newspaper articles. The means of the number and the absolute sum of salient onsets did not significantly differ between two groups: t(38)=0.1, p=0.91 , and t(38)=0.02 , p=0.98 , respectively.

	Phoneme				**Onset**				**Nucleus**				**Coda**			
	Num		Sum		Num		Sum		Num		Sum		Num		Sum	
Text	NEWS	POEM	NEWS	POEM	NEWS	POEM	NEWS	POEM	NEWS	POEM	NEWS	POEM	NEWS	POEM	NEWS	POEM
1	3	6	3.56	8.39	11	12	6.00	11	3	3	2.58	2.71	8	13	4.93	14.76
2	7	9	14.76	22.01	9	11	8.83	13.42	3	6	8.19	9.70	10	11	10.42	10.57
3	5	4	5.45	20.67	3	11	8.43	20.14	2	2	1.03	2.45	10	9	13.36	9.065
4	5	6	5.45	10.47	3	9	8.43	5.67	2	3	1.03	7.17	10	6	13.36	14.22
5	4	7	14.04	12.50	9	4	8.45	3.44	2	6	10.71	6.84	12	8	10.59	6.26
6	9	7	17.26	13.97	17	9	17.32	6.38	4	4	7.74	7.33	7	8	6.06	9.95
7	5	10	17.79	15.89	9	5	18.71	6.40	2	4	4.61	12.2	9	15	13.10	20.56
8	3	6	3.83	9.43	5	5	6.26	8.15	3	3	3.20	1.75	3	12	4.68	9.87
9	9	6	20.95	20.21	7	6	12.10	6.46	3	6	8.41	21.06	7	7	6.06	2.33
10	5	10	10.31	29.45	9	9	12.82	8.05	0	6	0.00	24.69	6	10	2.69	16.43
11	2	3	1.10	9.46	5	6	4.34	8.76	0	4	0.00	11.3	9	10	5.37	5.44
12	8	8	24.98	11.39	11	8	18.53	5.56	5	3	9.55	7.39	10	7	4.96	3.44
13	8	7	24.19	14.66	11	9	14.02	19.91	4	1	12.26	1.41	3	11	6.78	11.08
14	8	9	12.89	12.22	7	10	8.53	7.49	5	4	10.06	5.53	8	12	9.44	11.28
15	7	8	13.75	15.36	9	7	6.57	12.75	2	3	8.96	7.23	3	11	6.01	7.57
16	5	9	15.23	14.84	6	4	4.03	4.10	2	4	10.35	14.76	6	10	10.54	23.32
17	4	10	3.11	32.75	4	8	2.37	12.72	2	7	2.92	18.21	8	9	8.67	10.13
18	6	9	8.86	26.62	10	11	8.44	12.13	2	6	4.88	13.07	8	5	6.34	14.76
19	4	9	11.62	19.89	6	11	2.43	9.05	2	3	7.09	8.95	8	8	4.34	5.85
20	1	8	2.27	10.47	9	3	7.69	1.99	0	4	0.00	8.95	2	11	1.92	13.72

Table 10.3. Numeric representation of the sublexical salient units for all of 20 poems and newspaper texts. "Num" stands for the number of distinct phonological units being salient (the 1st indicator) and "Sum" for the absolute sum of all segments positioned outside the confidence interval (the 2nd indicator).

These results are consistent with the theory of foregrounding at the phonological level. All poems used for the present analyses contain shared rhymes across endings of lines. Considering the internal structure of a syllable, where rhymes involve syllabic nuclei and codas, the significant higher mean of the number and the sum of salient nuclei and salient codas in poems as compared to prose texts reflects how such typical features of lyrical texts using foregrounding at a phonological level can emerge in the analyses conducted with our tool. When selecting texts for the present exemplary analyses, we focused on rhymes as the one lyrical feature to be compared between lyrical texts at the one hand and more prosaic everyday language use at the other. Meeting our hypothesis, results from the tool showed that in particular for those sub-syllabic segments that are essential for rhymes, syllabic nuclei and codas, significant differences between the two groups of lyrical and non-lyrical texts were obtained.

But certainly, the use of other stylistic techniques and sound effects such as alliteration, consonance, assonance and in particular phonological iconicity in poetry or related literary genres, could obviously also be addressed by our tool once particular hypotheses have been formulated and adequate texts have been chosen for analyses. Results obtained for syllabic onsets from the poem Totenklage (see Figure 10.9) or the finding of significantly higher means of the number and the sum of salient phonemes in poems as compared to control texts might suffice as initial examples for the potential of such enterprises.

10.5 Discussion

In this paper, we presented a computer-linguistic tool automatically transforming digitally presented text into phonetic transcriptions, the absolute frequencies of which are documented and compared to expectation values derived from a large scale database of 25 Million German words representing everyday language use.

The user of this tool is provided with the information of whether any specific sublexical unit occurs more or less often than this would be expected for exactly this specific unit given the length of the text it was taken from and, importantly, the corresponding degree of derivation from expectation values is an object of a statistical inference process assigning levels of significance for deviations based on a probabilistic model.

Our tool, thus, lays the ground for all kinds of empirical research relying on the assumption that what is generally understood as higher level intentional message or emotional tone of a text may already be observable at the sublexical level. Beyond providing a text analysis tool for establishing frequencies of occurrences of phonologically defined sublexical units of different grain size within texts, the novel contribution of our tool mainly consists in implementing a mathematical model which offers a straightforward way to transform the mere observation of these frequencies of occurrence into inference statistical conclusions regarding their significance – thus, resolving the problem of how to define and assess the salience of small units of language embedded in large texts.

Once the frequency of occurrence of a given phonological sublexical unit within a text is identified by the presented tool as possessing a salient status, it is reasonable to assume that the specific unit might be used in a given text as an element of foregrounding – especially in the artistic domain of language use where the connection between formal aspect of language and higher meaning is of great importance.

It's worth noting that we opted for a phonological definition of these potentially relevant foregrounded elements in texts, because most literature on correspondences between meaning and sublexical structures traditionally focuses on language's sound as the relevant perception domain. In consequence, all analyses of the present tool are based on phonologically described sublexical units. Of course, this does not preclude an additional, and potentially independent, role of specific visual features of words or letters as foregrounded elements with potential iconic status (see for instance Doyle & Bottomley, 2009; 2011). A differential examination of phonological vs. visual phenomena is, yet, beyond the scope of the present paper.

The sublexical salience fingerprint of written texts – provided by our tool – can then also be related to theories and approaches claiming a specific emotional value for single sounds or other sublexical phonological units in language. Once their salience has been established, the emotional status of such sublexical units – potentially being used as foregrounding elements in the given context – could then serve to predict the sublexical emotional tone of a given text as a function of the emotional status of salient sublexical units.

Such an enterprise may be useful for different types of texts and different research motivations ranging from general to applied linguistics or psychology. Its use may though, be especially evident in the case of lyrical language: For instance, the emotional tone of the poem "Totenklage" (the first poem reported in this article) might be predicted by focusing on salient vowels in this text as following. The "frequency code" theory by John Ohala (1994; 1996) posits sound-meaning correlations in intonational communication of affect and in iconic vocabulary: whereas high pitch sounds signify smallness, non-threatening attitude, desire for goodwill of the receiver, and generally positive emotional valence, low pitch sounds convey largeness, threat, self-sufficiency, and generally negative emotional valence. Regarding this categorization, vowels with lower pitch (e.g., /u/, /o/, /a/) might be associated with negative emotion and vowels with higher pitch (e.g., /i/, /e/, /@/ and to some extent /6/) with positive emotion. Considering the extracted salient vowels in Hugo Ball's "Totenklage", the emotional tone of this poem at sublexical level (based on Ohala's theory) would be revealed as negative: where /o:/ and /a:/ appear significantly more than expected, the poem includes /@/ and /6/ significantly less than it was expected. This finding seems perfectly consistent with poet's presumable intent as evidenced by the poem's title "Totenklage" (lamentation of the dead) as the only clue to the potential semantic content of the poem.

Additionally to the analysis of literary texts and in a more general context, the presented approach in this article can be used as a complementary method for sentiment classification of written texts. Recent approaches on sentiment and content analysis have awarded much attention to the determination of the contextual polarity of written texts. To this end, different strategies of the Computer-Assisted Text Analysis (CATA) have been applied in an attempt to link word use to the affective state of a given text. These approaches, however, are mostly based on predefined keywords or clusters of keywords and lack a potentially important indicator of a text's emotional state; the sound. On this note, our presented approach can expand upon existing methods for sentiments classification and opinion mining.

In comparison to the analysis of affective and aesthetic effects at the *lexical* level, the *sublexical* level has received less attention by researchers. For instance, in recent years large scale databases for the emotion content of words have been collected (see, for instance, BAWL (Võ et al., 2009) for German, and ANEW (Bradley & Lang, 1999) for English), and emotion effects during visual word recognition have been reported using a wide range of methodologies: event related potentials (Conrad et al., 2011; Hofmann et al., 2009), hemodynamic responses (Kuchinke et al., 2005; 2006), pupillometry (Võ et al., 2008), and response times (e.g., Briesemeister et al., 2011a, 2011b; 2012; Kousta et el., 2009). Even more recently, also at a *supralexical* level of sentences or larger text passages, emotional and aesthetic effects during reading have been reported (Bohrn et al., 2012a, 2012b; 2013; Altmann et al., 2012a, b; Hsu, Conrad, and Jacobs, 2013). The current approach extends this line of research by the focus on such effects at the *sublexical* level. Of course, further research is needed to obtain empirical evidence for the existence of emotional and aesthetic effects of sublexical units at the level of functional neural correlates and substrates.

In concluding, we would like to stress that it is this practical solution of how to define and assess salience that might help theories and hypotheses concerning iconicity or general sound to meaning correspondences in language cross the gap between hermeneutic interpretation and principles of test and falsification widely separating different branches of science involved in the investigation of language and the arts.

Chapter 11

Measuring the basic affective tone of poems via phonological saliency and iconicity[8]

11.1 Abstract

We investigate the relation between general affective meaning and the use of particular phonological segments in poems, presenting a novel quantitative measure to assess the basic affective tone of a text based on foregrounded phonological units and their iconic affective properties. The novel method is applied to the volume of German poems “verteidigung der wölfe” (defense of the wolves) by Hans Magnus Enzensberger, who categorized these 57 poems as friendly, sad or spiteful. Our approach examines the relation between the phonological inventory of the texts to both the author’s affective categorization and readers’ perception of the poems – assessed by a survey study. Categorical comparisons of *basic affective tone* reveal significant differences between the three groups of poems in accordance with the labels given by the author as well as with the affective rating scores given by readers. Using multiple regression, we show our sublexical measures of *basic affective tone* to account for a considerable part of variance (9.5%-20%) of ratings on different emotion scales. We interpret this finding as evidence that the iconic properties of foregrounded phonological units contribute significantly to the poems’ emotional perception – potentially reflecting an intentional use of phonology by the author. Our

[8] This chapter is published as: Aryani, A., Kraxenberger, M., Ullrich, S., Jacobs, A. M., & Conrad, M. (2016). Measuring the basic affective tone of poems via phonological saliency and iconicity. *Psychology of Aesthetics, Creativity, and the Arts*, *10*(2), 191. http://dx.doi.org/10.1037/aca0000033

approach represents a first independent statistical quantification of the *basic affective tone* of texts.

11.2 Introduction

The discussion about an inherent relation between sound and meaning in language dates back to Greek antiquity with Plato's Cratylus dialogue (Plato, 1892). Despite such a long, and often controversial tradition (Genette, 1995), it is still an open question whether and to which extent the overall affective meaning of a text is (co-)determined by the specific use of sound in general, or clusters of phonological units in particular. With the present study we aim to contribute to answering this question, with a special focus on the literary genre of poetry. We will present quantitative phonological analyses of poems, the affective impact of which we assessed via a rating study.

In general, linguistics and literary studies have not paid much attention to the relationship between sound and meaning. Usually considered to be opposed to the linguistic principle of arbitrariness (Saussure, 1916; Hockett, 1958), research on this topic often fell short, or only reached dubious reputation due to methodological and theoretical shortcomings. It is just recently that the potential relation between sound and meaning received increasing interest. Several recent studies suggest a connection between formal aspects of language and meaning with phonological iconicity[9] as a general property that structures language in a supplementary way (Christiansen & Monaghan, in press; Perniss, Thompson and Vigliocco, 2010; Perniss & Vigliocco, 2014; see Schmidtke, Conrad, & Jacobs, 2014, for a review) and might be important for early language development (Monaghan, Shillcock, Christiansen & Kirby, 2014). At the lexical level, results of earlier empirical studies already had provided support for a systematic relationship between phonological and lexical features of words., e.g., attributes of conceptual meaning such as size (Sapir, 1929; Huang, Pratoomraj,

[9] To refer to the phenomenon of a sound-meaning relationship in this context, we use the term "phonological iconicity" (Aryani et al., 2013; Schmidtke et al., 2014), especially to focus on the relation between phonemes or clusters of phonemes and meaning.

Johnson, 1969, Thompson & Estes, 2011) or shape (Köhler, 1929; Ramachandran &Hubbard, 2001; Westbury, 2005), lexical category (Farmer, Christiansen, Monaghan, 2006; Fitneva, Christiansen, Monaghan, 2009), affective meaning (Ohala, 1994; Myers-Schulz, Pujara,Wolf, Koenigs, 2013) – including respective advantages for word learning (Nygaard, Cook, & Namy, 2009; Monaghan, Christiansen, and Fitneva, 2011).

Beyond the lexical level of language, and especially in relation to affective meaning and emotions (see Hsu, Jacobs, Citron, & Conrad, 2015a, for lexical affective effects during text reading), the literary genre of poetry seems to be of particular interest for the investigation of phonological iconicity (Albers, 2008; Aryani, Jacobs, Conrad, 2013; Auracher et al., 2011; Jakobson & Waugh, 1979/2002; Fónagy, 1961; Schrott & Jacobs, 2011; Tsur, 1992b). Poetry can generally be understood as inherently concerned with the expression and elicitation of emotions (Lüdtke, Meyer-Sickendieck, & Jacobs, 2014; Meyer-Sickendiek, 2011; Winko, 2003) while being deeply rooted at the aesthetic and perceptual level in the domains of speech and sound (Jacobs & Kinder, 2015; Schrott & Jacobs, 2011; Wolf, 2005). Emphasis on phonological units such as syllables or phonemes through diverse stylistic devices, like onomatopoeia or figures of self-similarity and parallelisms as rhyme, meter or alliterations may serve as examples for the latter while the presence of meter and rhyme, for instance, has been shown to affect aesthetic appreciation, intensity of processing and emotional perception of lyrics and poems (Menninghaus et al., 2014; Obermeier et al., 2013, Bohrn et al., 2012 b; 2013).

Considering two major principles of the poetic genre, i.e., the prominence of sound properties and expressed or perceived emotions, several empirical studies offer first affirmative evidence for a relation between affective meaning and occurrence of specific phonological units in poetry. In a comparative analysis of Old Egyptian hymns and lamentations together with hymns and ballads by Johann Wolfgang von Goethe, plosive sounds were found to occur significantly more frequently in hymns of both sources, whereas nasals were more frequent in lamentations and ballads (Albers, 2008). Similarly, cross-linguistic studies report a higher frequency of the plosives /p/,

/b/, /t/, and /d/ in poems rated as happy vs. a higher frequency of nasals in poems perceived as sad; with consistent results for German, Chinese, Russian and Ukrainian participants and poems (Auracher et al., 2011). Likewise, German and Brazilian participants found plosives to be more appropriate in a pleasant context (for instance a wedding) than the nasals /m/, and /n/, which seemed to be more suitable to express sad feelings (for instance in the context of a funeral) (Wiseman & van Peer, 2003). Some other studies draw exclusively on the works of a single author: Miall (2001) compared passages from Milton's "Paradise Lost" that either dealt with depictions of Hell or Eden. Passages about Hell were found to contain significantly more front vowels and hard consonants than passages about Eden while the latter contained more medium back vowels. Analyzing the phonological material of the poetic works of Edgar Allan Poe, Whissell (2011) reported that Poe used "pleasant, sad, and soft sounds" more frequently than sounds that were categorized as "active". Note that these results are based on inductive phonoemotional classification, i.e., on the tendency of single phonemes to appear more often or more seldom in English words with known emotional meaning, derived from a rating study (Whissell, 2000).

However, despite these preliminary indications of a relation between sound and meaning in poetry—and of fundamental interest for our present undertaking—, no systematic empirical research so far has offered an independent statistical measure to quantify the affective tone of texts or poems based on the given phonemic material. So far, research merely observed the presence or absence of single phonemes or classes of phonemes, resulting in statistical reports of categorical distinctions regarding the phonemic inventory of different poems. Furthermore, taken together, most studies arrived at contradictory results concerning phonemes under observation as well as the affective meaning associated with them. For instance, Wiseman and van Peer (2003), Albers (2008), and Auracher et al. (2011) assigned the plosives /t/, /b/, /d/ and /p/ to the affective category of happiness, Fónagy (1961) attributed /t/ to aggressive poems, while according to Whissell (1999) plosives like /t/, /b/, and /d/ tend to be more dominant in unpleasant words and stand in a negative correlation with pleasantness.

11.2.1 The Present Study

In the following, we focus on three types of theoretical and methodological limitations and shortcomings of previous studies while presenting a new approach to overcome these deficiencies concerning:

i) an adequate statistical operationalization of affective sound at the sublexical level,

ii) a theoretical framework regarding literary communication (considering the emotional classification of poems by both readers and author), and

iii) the varying and often insufficient operational definitions of emotion and affective meaning.

To that end, we formulate an interdisciplinary framework that draws on literary theory, (psycho-)linguistics as well as psychology of emotions, and aims to develop a novel statistical measurement quantifying the *basic affective tone* of a poem.

(i) Previously used operational definitions of the sound component within an assumed relation between sound and meaning appear insufficient. Among other things, this may be responsible for discrepancies of results of previous studies mentioned above. To date, most available research concentrated on the mere or rational frequencies of occurrence of single phonemes or classes of phonemes. This seems appropriate when the attribution of a poem to certain binary emotional categories is used as the independent, and the phonological material as the dependent variable. An example for this procedure is Albers (2008), where an ascription to the emotional categories of sadness and happiness was based on the poetic form (ballad vs. hymn) and on the content of every single poetic line while the variance of phonemic material (in this case the frequencies of nasals vs. plosives) was treated as dependent variable. However, such a strategy might face general problems in detecting systematic signals, because the specificity of poetic language may alter the general distribution of the phonological data to be analyzed. As we will argue, not the absolute frequency of occurrence of a certain phonological unit within a poem, but rather its deviant occurrence –compared to prosaic language— might be most relevant to capture the *basic affective tone* of a poem. This represents an important methodological issue that has, to our knowledge, not been considered before. Our focus on deviant elements within a text is justified by the notion of a de-automatization of the reading process, often called 'foregrounding' (Garvin, 1964; Jacobs, 2015a, b; Lüdtke et al., 2014; Miall & Kuiken, 1994; van Peer, 1986; Mukarovsky, 1932/1964; see Hakemulder, 2004 for a short overview; and Sanford & Emmott, 2012 for a broader discussion on the topic). Foregrounding refers to the stylistic device of defamiliarization as well as the general deviation between prosaic and poetic language (Shklovsky, 1925/1990; Mukarovsky, 1964, Jakobson, 1960). Our explicit consideration of foregrounding draws on the notion of figure-ground elements in Gestalt psychology (Rubin, 1921; van Holt & Groeben, 2005), standing in line with the assumption that deviation within literary texts always refers and relates to the standard concerning linguistic rules and norms, or literary conventions and canons (Iser, 1976/1994). Differences between poetic and prosaic language use can be based on the dominance of the poetic function including the focus within the message on the message itself (Jakobson, 1960). Consequently, this determining function of the poetic genre influences all linguistic constituents of poetry and particularly its sound (Jakobson & Waugh, 2002), i.e., the phonological structures and their units; phonemes and syllables.

Empirical results support the notion that literary foregrounded and hence deviating elements provoke a more intensive and extensive cognitive processing (van Peer, 1986) and deeper emotional experience (Miall and Kuiken, 1994). Furthermore, at the level of functional neural correlates and substrates, Bohrn et al. (2012b) found that foregrounding leads to an enhanced activation in affect-related regions (orbitofrontal cortex, amygdala) and also increases cognitive processing demands (see also Hsu, Jacobs, Altmann, & Conrad, 2015b).

Therefore, in the present study, we focus on salient phonological units as a potential source influencing readers' affective perception of a poem. To extract such salient units, potentially used as foregrounded elements, we use the statistical model developed and validated by Aryani et al., (2013). This model compares the frequency of occurrence of a phonological unit in a given text with an expected value based on a probabilistic model to assess deviations from standard language use (see Material and Methods for details).

A further methodological advancement of our approach is that we capture the *basic affective tone* in an independent and in a quantitative manner; based on a novel method focusing on apparent general sound to meaning correspondences within the German language. This method enables us to predict the affective load of phonological structures in a single poem without the necessity of further comparisons. For this, we utilize previous findings on the relationship between the meanings of single words at the micro- and macro-level. We extended related approaches (Heise, 1966) and defined a *sublexical affective value* for each syllabic unit (i.e., onset, nucleus, or coda) in the German language. These values are based on the average ratings of emotional valence and arousal of words in which a certain syllabic unit occurs. Importantly, to determine the *basic affective tone* of a text, we focus on those syllabic units (and their *sublexical affective values*) that can be considered foregrounded elements.

(ii) Another critical point of present approaches to phonological iconicity is that research on the contribution of phonological features to affective meaning in poetry lacks an inclusive theoretical model that would allow for considering not only the poems, but also readers and author and their emotional classification of poems. Hitherto, research has either concentrated solely on the textual constituents (e.g., Albers, 2008) or put these in relation with results from survey studies (e.g., Auracher et al, 2011). Only a few studies have also paid attention to the poets –after all the creators of their stimuli—and where this was the case, only theoretical treatises or critical writings were considered (e.g., Whissell, 2002; 2011).

To surmount these critical points within an extensive approach, we use Jakobson's model of language function (Jakobson, 1960)—an extended version of Bühler's (1934) organon model—as theoretical starting point guiding our argumentation and methodological procedure. In particular, we refer to the basic constitutive factors of communication (i.e., the addresser, the addressee, and the self-referential message) that can also be seen as main contributors to literary communication (Figure 11.1). In extension of Jakobson's model, and in accordance with the Panksepp-Jakobson hypothesis of the Neurocognitive Poetics Model (NCPM; Jacobs, 2015a,b; Jacobs et al., 2015), we further consider these in regard to the encoding and decoding of emotions in texts (Kraxenberger, 2014).

The following two main considerations were crucial for the selection of poems we used as stimuli. First, the author's emotional classification of his own poetic works would provide an independent variable that would—also statistically—"empower" our investigation. Second, we decided to work with contemporary poems; underrepresented or missing in previous empirical research on phonological iconicity (Schmidtke et al., 2014).

The chosen volume, "verteidigung der wölfe" (defense of the wolves) by the German poet Hans Magnus Enzensberger[10], first published in 1957, consists of 57 German poems that are mainly written in unrhymed, free verse[11]. Enzensberger's poems represent a compliant and suitable set of stimuli not only because of the temporal proximity between text and readers, but also because all poems of "verteidigung der wölfe" are categorized by the poet himself into three distinctive affective categories: friendly poems (*freundliche Gedichte*), sad poems (*traurige Gedichte*) and spiteful (or angry) poems (*böse Gedichte*). This affective categorization of the poems by the author is not understood to be sarcastic or cynical (Walser, 1999). Note that these affective categories are of considerable importance, notably for literary studies, since the affective categorization of the poems by Enzensberger allows drawing on classical rhetoric and aesthetics, especially in regard to Friedrich Schiller's poetic tripartite of elegy, idyll and satire (Schiller, 1795/1981; Grimm, 1981). This leads ultimately to a functional distinguishability of the respective poems and in particular when considered in terms of generic categorization as "kinds of emotions" (Watanabe-O'Kelly, 1981).

[10] Enzensberger, born in 1929, is praised to be one of Germany's most important poets (Astley, 2006). He published numerous volumes of poems which have been translated into several languages.
[11] Within the German poetic tradition, one would rather speak of "unbound verses" (*ungebundene Verse*), which are typical for the 20th century and are neither rhymed, nor follow antique strophic forms.

(iii) Most studies on phonological iconicity vary in terms of their interpretative construction as well as operationalization of emotion or affective meaning in text. In our understanding, a text or a poem implies a "general affective meaning" that is inscribed at all levels of language, including the phonological one. Herewith, we are not referring to fine-grained, possibly altering affective experiences while reading a poem, which can rely on both, dynamic shifts in its plot and depiction, or subjective reading experience. Rather, the general affective meaning denotes the overall theme of a text and constitutes the emotional meta-perception of the poem in form of appraisals, i.e., the perception of the poems by their readers as rather positive or negative, rather arousing or calming, etc., hence representing a principle of affective attribution by the reader (Lüdtke et al., 2014). For the present study, we tried to capture these attributions of the reader via semantic differentials (Osgood 1952, Osgood, Suci, & Tannenbaum, 1957) in a rating study (see Material and Methods for details) on the dimensions of valence and arousal. These affective dimensions allocate emotions in a bi-dimensional space and are used by influential dimensional models of affect (e.g., Wundt, 1896; Russell, 1980; Bradley and Lange, 1999; Recio et al., 2014). Because of its applicability to expressed and perceived emotions, or psychological construction in general (Russell, 2009), we used this approach to operationalize the general affective meaning of the poems in our study (see also Auracher, 2011). Our operationalization and measurement of the affective load of phonological structures, the *basic affective tone* that is understood to potentially reflect a part of the general affective meaning at the phonological level, is, accordingly based on the same dimensional approach.

To sum up our intentions: With this study we present a novel quantitative measure of the *basic affective tone* of poems. We hypothesize that this measure can capture significant sublexical contributions to the general affective meaning of a poem, and, in consequence, that sound in poetry can contain a semantic, affective function reflecting the poet's intentions and influencing readers' perception of the general affective meaning of a poem (see Figure 11.1).

11.3 Material and Methods

In what follows, we first describe the method for measuring the *basic affective tone*. The method consists of 1) calculation of *sublexical affective values* (*SAVs*) for each syllabic unit in German language, 2) extraction of foregrounded phonological units from each poem –or any digitally available text, in general—based on a probabilistic model developed in a previous study (Aryani, et al., 2013), and 3) quantification and statistical evaluation of the *basic affective tone* of the text submitted to the model based on the previous steps.

Further, we will describe the procedure of our rating survey assessing the "general affective meaning" of the poems.

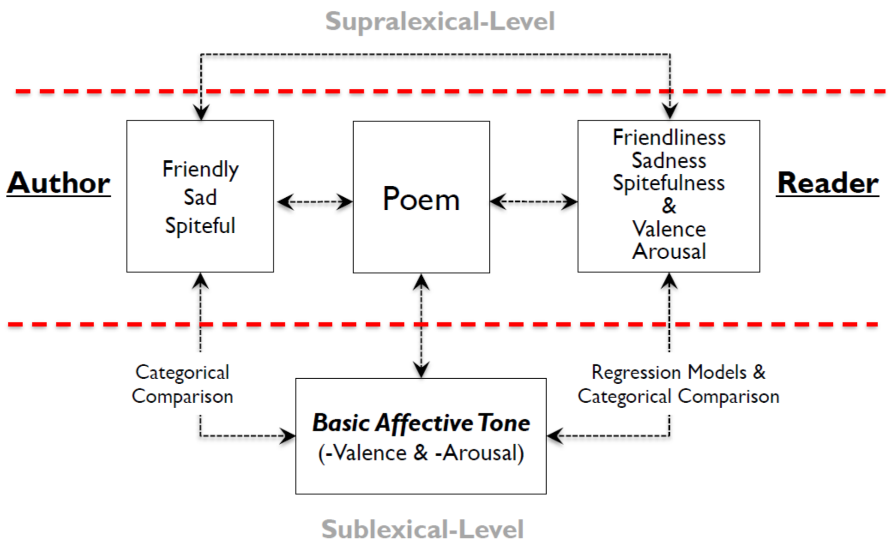


Figure 11.1. The detailed procedure of analysis of poems based on Jakobson's model of language function, i.e., the communication between sender (the author, left side) and receiver (the reader, right side) through the message (poem, center). The ascription of a certain theme (affective meaning) to a text, understood as a form of meta-perception, is formed at both the supralexical (top) and the sublexical (bottom) level of language. Our measurement of the *basic affective tone* is purposed to capture the latter.

11.3.1 Sublexical Affective Values (SAVs) of Syllabic Units

To investigate the potential emotional load of single phonological units, we pursued a systematic analysis of a normative database comprising 5300 German words. This database extends the first normative database providing ratings on emotion scales for German words (BAWL: Võ, Jacobs, Conrad, 2006; Võ et al., 2009) that has been validated in numerous behavioral and neuroscientific studies (Conrad, Recio, & Jacobs, 2011; Jacobs et al., 2015; Kuchinke et al., 2005). The word entries therein were rated by at least 20 subjects for a) emotional valence varying from negative (-3) to positive (+3), and b) emotional arousal varying from low (+1) to high (+5) based on a dimensional model of affect (Russell, 1980; Bradley and Lang 1999; Wundt, 1896).

Based on apparent general sound to meaning correspondences within a language, and by following the idea that affective meaning of language units (at the macro-level) co-varies with phonological structure (at the micro-level) (Lamb, 1964; Heise 1966), we encoded all words in the database according to presence or absence of certain phonological units to calculate a *sublexical affective value* (*SAV*) for each unit.

For the following reasons, we opted for syllabic units (i.e., onsets, nuclei, and codas) instead of single phonemes as presumably most effective sublexical units regarding *sublexical affective values* (*SAVs*): Research in psychoacoustics has shown that iconic characteristics of sound may be bound to a different linguistic level than the one of single phonemes, e.g., to the dynamic shift within words' first two frequency components (Myers-Schulz et al., 2013), fundamental frequency (Bänziger & Scherer, 2005), and spectral center of gravity (Sauter, Eisner, Calder, & Scott, 2010). As boundaries between syllables often mark interruptions of the ongoing stream of speech within words, syllabic structure offers a most basic segmentation device for phonological word forms, and several empirical reports for different languages have shown that phonological syllabic units serve as functional units of language processing even during silent reading (see, e.g., Conrad & Jacobs, 2004; Conrad, Grainger, & Jacobs, 2007; Conrad, Carreiras, Tamm, & Jacobs, 2009).

Accordingly, and also due to the contradictory results reported in previous studies (see introduction), rather than using single phonemes, we calculated *SAVs* for more complex sound clusters arising from a syllabic definition of phonological units, i.e., syllabic onsets, nuclei and codas. Note that these syllabic components are the most effective sublexical units mediating visual word recognition (Nuerk, Rey, & Jacobs, 2000) based on which orthographic-phonological information is organized during the reading process (Jacobs, Rey, Ziegler, & Grainger, 1998).

Therefore, we segmented all 5300 words into syllabic units (in phonological form). For the calculation of the *SAV* of each syllabic unit, we considered all words in the database, in which the syllabic unit appeared at least once in every word. Valence and arousal values for each single syllabic unit (Figure 11.2-A) were calculated averaging emotional valence and arousal ratings of all those words a syllabic unit occurred in. Resulting values were then standardized with respect to the calculated values of all existing units and their frequency of occurrence in the words list. These standardized values were then assigned to corresponding syllabic units as their *SAVs* for both affective dimensions of valence and arousal.

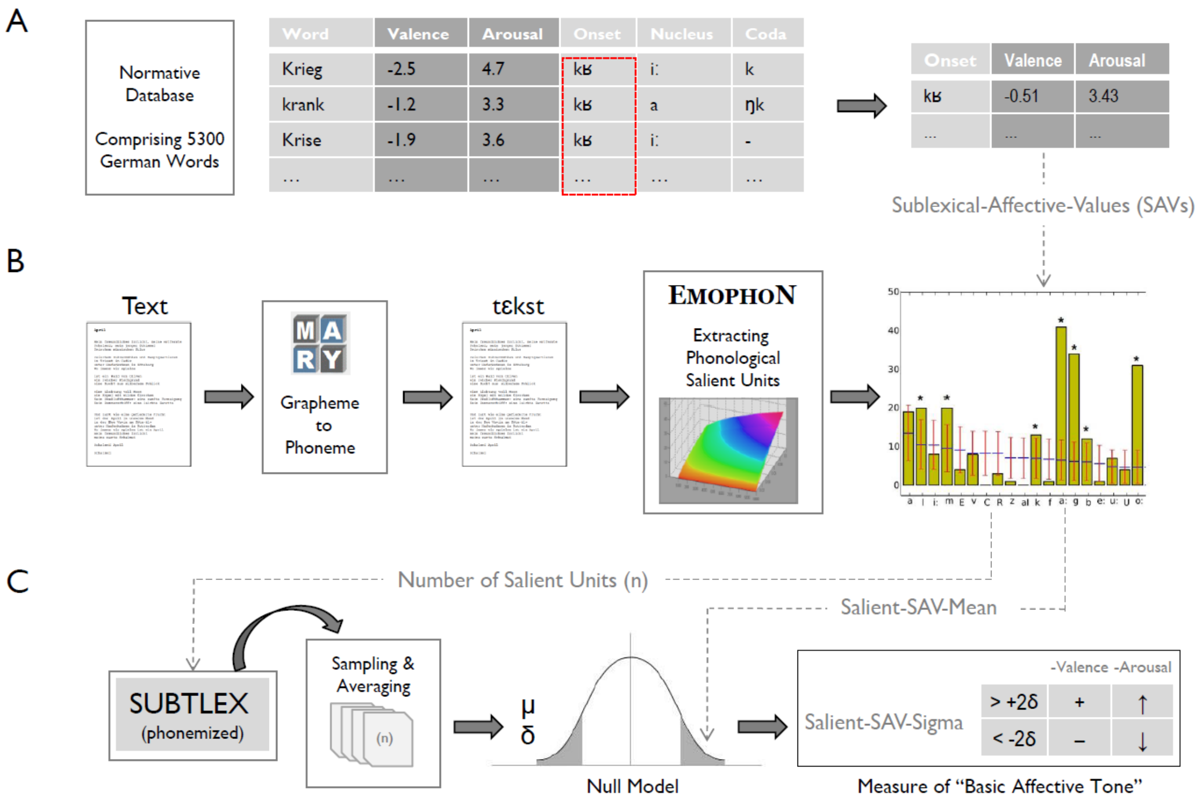


Figure 11.2. Calculation of sublexical measures of the *basic affective tone* for a given text. **A)** *sublexical affective values* (SAVs) of all syllabic units are calculated based on the average ratings of words containing a certain syllabic unit (see example of /kʁ/) **B)** A given text is phonemized using the G2P-software MARY and its salient syllabic units are subsequently extracted via a probabilistic model integrated in the "EMOPHON" **C)** The *basic affective tone* of the text is calculated based on the mean of SAVs for salient units. This mean value (Salient-SAV-Mean) is compared against an exhaustive distribution of random samples with matching numbers of units, to test for significance of deviations concerning SAVs – finally represented by *Salient-SAV-Sigma*.

11.3.2 Extraction of salient phonological units

As the *basic affective tone* is directly dependent on phonological features of texts, each poem to be analyzed was converted into phonetic notation by using the German text-to-speech synthesis system MARY (Schröder & Trouvain, 2003). To phonemize the text automatically, MARY works with an extensive lexicon dealing with known words and a letter-to-sound conversion algorithm dealing with unknown words. All 57 texts—in a phonemized form—were screened for salient phonological segments using EMOPHON (Aryani et al. 2013; see Figure 11.2-B). EMOPHON's measure of phonological salience is based on the deviation of the observed frequency of occurrence of particular phonological syllabic units in a given text from their respective expected frequencies: salient units, potentially being used as foregrounded elements, are those occurring significantly more frequent than could be expected based on the SUBTLEX-DE linguistic corpus (Brysbaert et al., 2011) – a database that presumably best represents prosaic, everyday speech. The exact method of detecting salient units is described in detail by Aryani et al. (2013).

11.3.3 Basic affective tone

To quantify—and further submit to inferential statistical tests—the *basic affective tone* of a text, we focused on its salient syllabic units and combined them with their respective *sublexical affective values* (*SAV*s). We first calculated a weighted mean value of *SAV*s of these salient units - used over-proportionally in a given text. For instance, if the syllabic units $s_1, s_2,\ldots,s_n$ are detected by EMOPHON as salient, the corresponding mean value of *SAVs* of these salient units (henceforth: *Salient-SAV-Mean*) for this text –in the case of arousal—would be:

$$\textit{Salient-SAV-Mean (arousal)} = \frac{\sum_{i=1}^{n} \mathrm{num(si)} \times SAV\mathrm{,aro(si)}}{\sum_{i=1}^{n} \mathrm{num(si)}}$$

Where num(s_i) is the number of a specific salient syllabic unit in the whole text, and *SAV*,aro(s_i) is the corresponding *SAV* (for arousal) of this unit.

In order to set a frame for interpretation of specific results, we test each *Salient-SAV-Mean* value for a given text against a null model. For this, we empirically calculated the distribution of the null model as a function of the number of salient units in each text. We therefore randomly pulled numerous chunks of syllabic units – with the same length as the sum of all salient units in a text—from the same corpus used in EMOPHON (i.e., SUBTLEX-DE). For each random sample, the mean of *SAVs* of the including syllabic units (for both valence and arousal) was calculated. We repeated this sampling-and-averaging process, with replacement, for 1 Million times to ensure that the number of samples is representative for a larger population of syllabic units and to obtain a good proxy of the null model (a procedure similar to the Mantel test). Based on the acquired data, the corresponding mean (μ) and standard deviation (δ) of the average *SAVs* of all samples (1 Million measures) were then calculated for each text (Figure 11.2-C). These acquired values of mean and standard deviation of the null model provide the possibility to interpret every specific value of the *Salient-SAV-Mean* of each text. Note that the overall mean values of the null models (for each text) were very close to nil because of the previous standardization of *SAVs*. We next divided the *Salient-SAV-Mean* of each text by the corresponding standard deviation of the null model, calculated in the prior step, thereby obtaining two parameters (valence, arousal), each of which indicates the distance of the respective mean value of SAVs of salient units (*Salient-SAV-Mean*) from nil in the form of $n\times\delta$. When comparing these parameters with the conventional confidence interval defined as $2\times\delta$ [12], it can be seen at a glance whether, and to what extent, the factor *Salient-SAV-Mean* deviate from an expected value determined by the null model. We refer to these novel parameters as *Salient-SAV-Sigma* (*-valence*, *-arousal*) representing statistical measures of the *basic affective tone* of a text at the sublexical level.

Since the *SAVs* of each syllabic unit are based on emotion ratings of the words comprising these units, our *Salient-SAV-Sigma* statistical measures may correlate – at least to some degree - with the mean affective values of the words composing a text.

[12] The confidence interval is originally defined as $\mu \pm 2\times\delta$, where μ stands for the mean and is equal to zero in this case.

To eliminate potential resulting problems of circularity, we calculated two additional control measures to control for this possible confound in all further analyses: For this purpose, we repeated the whole procedure for each poem, using this time ALL syllabic units of each poem - rather than only the salient units extracted by EMOPHON. The resulting control measures, named *Control-SAV-Sigma* (one for valence and one for arousal), based on the mean of *SAVs* of all syllabic units in a text, should help clarifying whether any potential effect of our *Salient-SAV-Sigma* measures would also be given without referring to phonological salience – or might have been driven by lexical affective values of words contained in the text. In the following group- and regression-analyses of the poems, these two control measures (i.e., *Control-SAV-Sigma-valence* & *-arousal*) are used as checkups for the two *Salient-SAV-Sigmas* which we use to operationalize the *basic affective tone* of texts. We then attempt to predict the general affective meaning—assessed by subjective ratings—as well as the author-based categories

11.3.4 Rating of the poems on emotion and affective dimensions

To assess the perceived emotionality of poems by the reader, we conducted an online survey for subjective ratings of each poem's content (i.e., the general affective meaning or overall theme) in regard to 1) affective valence and arousal (matching our sublexical dimensions), and 2) the three emotion categories created by the author. We asked participants to give their subjective rating on the following affective scales: I. *valence*: on a 7-point rating-scale ranging from -3 (very negative) to +3 (very positive), II. *arousal*: on a 5-point rating-scale ranging from +1 (very calming) to +5 (very arousing)[13], III. *friendliness*: on a 5-point rating-scale ranging from +1 (not friendly at all) to +5 (very friendly), IV. *sadness*: on a 5-point rating-scale ranging from +1 (not sad at all) to +5 (very sad) and V. *spitefulness*: on a 5-point rating-scale ranging from +1 (not spiteful at all) to +5 (very spiteful).

[13] To measure arousal, we combined verbal anchors with a non-verbal pictorial assessment, i.e., SAM (Bradley & Lang, 1994)

A total number of 252 German native speakers (173 female) between the ages of 17 and 76 years (M = 35.9, SD = 12.1) participated in this study[14]. Each poem was rated on average by more than 17 participants on each scale (min = 15, SD = 1.7). All poems were presented pseudo-randomly. Poems that were familiar to participants were excluded from the individual survey to eliminate the mere-exposure effect (Zajonc, 1968; 2001).

11.4 Results

In the following we report our results considering the main components of literary communication according to our extension of Jakobson's model of language functions (see Figure 11.1). Analyses address relations between emotional evaluations of the readers at the one hand, and of the author Enzensberger at the other, as well as potential contributions/correlations of the *basic affective tone*, derived from potentially foregrounded and affect-loaded phonological units, to such evaluations. In doing so, we first focus on the relation between author and reader in regard to their classification/evaluation of the poems, followed by analyses of respective relations with the textual, sublexical measures of *basic affective tone* (i.e., *Salient-SAV-Sigma*). Technically, categorical comparisons (used for author-related as well as reader-related analyses) will be combined with regression models (reader-related analyses).

11.4.1 Ratings of the Poems

Mutual linear correlations between ratings on different dimensions are displayed in Table 11.1. As to the classical general dimensions of the bi-dimensional affective space, the rating scores for valence and arousal are correlated negatively (r = -.7) indicating that the more negative a poem, the more arousing it tends to be – which is in line with a recent account of the general relation between the two affective dimensions for German words (Schmidtke et al., 2014). But valence and arousal ratings also showed tight correlations with ratings for the more specific emotion

[14] For our rating study we used a more recent edition of "verteidigung der wölfe" which, in contrast to the first edition, is not exclusively written in lower-case letters but represents the standard German orthography.

ratings: As could be expected, spitefulness ratings increased with arousal (r = .73) but decreased with valence ratings (r = -.77), whereas friendliness ratings displayed the opposite relation to the two affective space dimensions (r = -.73 for arousal and r = .89 for valence). Rating scores for sadness were tightly but inversely correlated with valence ratings (r = -.62), whereas only a rather weak correlation with arousal was given (r = .38) suggesting sadness to be an emotion with a somewhat fuzzy positioning in the affective space – presumably involving both calm and excited features.

Affective Categories						Dimensions	r				
Friendly (N=19)		Sad (N=21)		Spiteful (N=17)			Arousal	Valence	Spitefulness	friendliness	sadness
Mean	SD	Mean	SD	Mean	SD						
3.21	.32	3.41	.32	3.64	.17	Arousal		-.70	.73	-.73	.38
-.32	.74	-1.06	.52	-1.37	.5	Valence			-.77	.89	-.62
1.88	.41	2.11	.63	2.71	.45	Spitefulness				-.74	.33
2.09	.61	1.58	.45	1.40	.26	Friendliness					-.45
2.00	.35	2.63	.3	2.51	.39	Sadness					

Table 11.1. Means and standard deviations of ratings on the affective dimensions arousal, valence, spitefulness, friendliness and sadness for poems from three different author-based categories (left). Pearson Product-Moment Correlation Coefficients (r) between rating scores of the five affective dimensions (right).

11.4.2 Sad/Friendly/Spiteful Categories and Ratings (Author-Recipient)

We next used the author's tripartite classification of the poems as a categorical independent variable and ran analyses of variance (ANOVAs) to analyze differences of subjective ratings between the three groups. Poems were originally categorized by the author as spiteful (n = 17), friendly (n = 19), and sad (n = 21). Means and standard deviations of ratings for the different categories are shown in Table 11.1 (see also Figure 11.3-A, top). ANOVAs revealed significant effects of the author-based affective category on each of the rating variables, $Fs_{(2,54)} > 10.6$, $ps < .001$. Post-hoc comparisons revealed that the differences in the rating scores were almost always consistent with the intended categorization by the author: the friendly poems (as categorized by the author) were rated significantly friendlier than the sad and the spiteful poems, $t(38) = 3.45$, $p < .001$, $t(34) = 4.42$, $p < .001$, respectively. The sad poems were rated as significantly sadder than the friendly poems, $t(38) = 5.67$, $p <$ .001, but not than the spiteful poems. The rating scores for spitefulness were significantly higher for the spiteful poems than for the friendly, $t(35) = 4.82$, $p <$ .001, and the sad poems, $t(37) = 3.53$, $p < .001$.

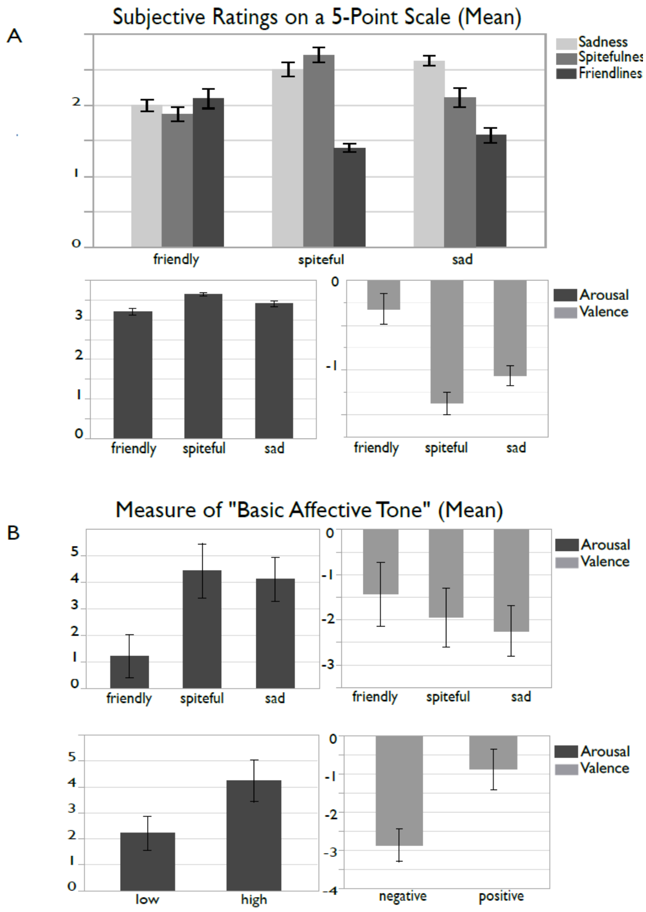


Figure 11.3. A) Rating scores on five affective dimensions for three author-based categories **B)** The *basic affective tone* of arousal and valence as measured by *Salient-SAV-Sigma* for the author-based categories (top) as well as the reader-based categories of valence and arousal (bottom). The Y-axis represents the mean of *Salient-SAV-Sigmas* for each category.

The categorical analyses also reproduced the affective author-based categories in terms of valence and arousal (Figure 11.3-A, bottom): Friendly poems were rated more positive in valence than sad, $t(38) = 3.92$, $p < .001$, and spiteful poems, $t(37) = 5.26$, $p < .0001$. The two latter categories did not significantly differ in valence. Regarding arousal, there was a significant effect of the poem's category, with spiteful poems being rated as more arousing than sad, $t(36) = 2.52$, $p = .014$ or friendly poems $t(34) = 4.6$, $p < .0001$, and sad poems more arousing than friendly ones, $t(38) = 2.25$, $p = .028$.

11.4.3 Basic Affective Tone of the Poems

In order to test how the above described author categories and emotion ratings relate to the sublexical phonological level, we conducted various analyses on our measurements of the *basic affective tone* of these poems (i.e., *Salient-SAV-Sigmas*). First, we compared respective values across the three author-based affective categories (Figure 11.1, bottom-left) as well as each time two reader-based categories of valence and arousal. Second, we attempted to predict the rating scores for these poems via regression analyses and categorical comparisons (Figure 11.1, bottom-right) in order to explore whether and to which degree our sublexical measures of the *basic affective tone* could significantly explore variance of, or predict ratings on emotion or affective scales.

11.4.4 Sad/Friendly/Spiteful Categories and Basic Affective Tone (Author-Text)

ANOVA results on *Salient-SAV-Sigmas* for poems from different categories assigned by the author (see Figure 11.3-B, top) showed a significant effect of category on the sublexical measure of arousal: $F_{(2,54)} = 4.04$, $p = .02$, but neither on that of valence, nor on any of the control measures that were used to control for potential circularity or a confound between lexical and sublexical values: $1.4 > Fs_{(2,54)} > 0.4$. The post-hoc comparisons of the sublexical measure of arousal indicated significantly higher means for the spiteful and sad poems compared to friendly poems:

spiteful>friendly, $t(34) = 2.52$, $p = .014$, sad > friendly, $t(38) = 2.4$, $p = .02$, but not between spiteful and sad ones, $t(36) = .25$, $p = .8$.

To further test whether phonological salience –as operationalized by the EMOPHON —was a crucial factor underlying these results, we conducted a One-way ANCOVA with category as between-subjects factor (sad, friendly and spiteful) and *Control-SAV-Sigmas* (control measures based on all rather than only on salient phonological units) as covariates. Similar to the ANOVA-results, there was a significant effect of category on the sublexical measure of arousal (*Salient-SAV-Sigma-Arousal*) after controlling for *Control-SAV-Sigma-Arousal*, $F_{(2, 54)} = 18.6$, $p < .0001$, but not on the sublexical measure of valence (*Salient-SAV-Sigma-Valence*) after controlling for *Control-SAV-Sigma-Valence*, $F_{(2, 54)} = .98$, $p = .38$.

11.4.5 Valence and Arousal Categories and *Basic Affective Tone* (Text-Recipient)

In addition to the preceding author-based categorization, we defined two new categorical variables based on median splits concerning the subjective ratings for affective valence and arousal. Thus, we divided the poems twice into each two groups with two levels: valence ("positive" vs. "negative") and arousal ("high" vs. "low") with almost the same number of poems in each category (valence: 28 positive vs. 29 negative; arousal: 29 high vs. 28 low). These comparisons should provide optimal contrasts for the exploration of relations between sublexical measures of the *basic affective tone* and ratings of affective impact of each poems as a whole – as independent and dependent variables relate to the same dimension.

To examine these relations, we conducted several one-tailed T-tests testing the following predictions concerning two sublexical measures (together with two control measures).: i.e., we expect a higher valence of *basic affective tone* for the "positive" as compared to the "negative" group, and higher arousal for the "high" vs. the "low" group.

T-tests on two sublexical measures (i.e., *Salient-SAV-Sigma-Arousal, Salient-SAV-Sigma-Valence*), as well as the two control measures (i.e., *Control-SAV-Sigma-Arousal*, *Control-SAV-Sigma-Valence*) revealed effects for our measures of the *basic affective tone* based on salient syllabic units, but not for the control measures (Figure 11.3-B, bottom). The sublexical predictor of valence (*Salient-SAV-Sigma-Valence*) was significantly higher in the "positive" group than in the "negative" group, $t(55) = 2.92$, $p = .002$. A categorical comparison between the arousal levels reveals a similar pattern: the sublexical measure of arousal (*Salient-SAV-Sigma-Arousal*) was significantly higher in the "high" group than in the "low" group, $t(55) = 1.93$, $p = .029$. No significant differences were detected for either the control measure of sublexical arousal or of valence (*Control-SAV-Sigma-Arousal* and *–Valence*) when comparing between the two respective levels of these dimensions: $t(55) = .95$, $p = .34$, for arousal, and $t(55) = .37$, $p = .71$, for valence.

Again, we conducted a One-way ANCOVA with category as between-subjects factor (valence: positive vs. negative, arousal: high vs. low) and *Control-SAV-Sigmas* as covariates to control for their potential contribution to effects of the sublexical measures of *basic affective tone* (*Salient-SAV-Sigmas*). Results revealed, again, similar main effects of category for valence, "positive" > "negative", $F_{(1, 55)} = 3.72$, $p = .0002$ (one-tailed) as well as for arousal, "high" > "low", $F_{(1, 55)} = 1.69$, $p = .047$ (one-tailed) on the corresponding *Salient-SAV-Sigmas,* this time after controlling for *Control-SAV-Sigmas*.

These results indicate that distributions of perceived valence and arousal of poems at the whole text level mirror our sublexical measures of the *basic affective tone* (Figure 11.3-B, bottom).

11.4.6 Multiple Regression Analyses (Text-Recipient, further evidence)

We performed several multiple-regression analyses exploring how affective qualities of particular salient phonological units, as reflected in our sublexical measuresof the *basic affective tone*, correlate with rating scores of each of the five emotion scales (Figure 11.1, bottom-right). These analyses reveal which, if any, sublexical measure significantly predicts participants' ratings on each of these emotion and affective scales. We used forward stepwise multiple regression with the minimum corrected AIC (Akaike information criterion) as stopping rule. This method appears a good choice for a screening procedure aiming to identify the most influential predictor among a set of competing intercorrelated predictors – leaving only residual variance to be explained by additional predictors after the strongest one has entered the regression model. The results (see Table 11.2) confirm that the sublexical measures, i.e., measures of the *basic affective tone* based on salient phonological units, account for a considerable part of variance of ratings in all affective dimensions; i.e., from 9.5% for sadness to 22% for spitefulness. The sublexical measure of arousal (*Salient-SAV-Sigma-Arousal*) appeared as the best measure of *basic affective tone* with the best potential to predict affective impact, as it was the sole significant predictor of emotion rating scores in four out of five models. Importantly, the sublexical control measures did not reach significance in any of the models (except the one for arousal, see next paragraph) which stresses the importance of phonological salience for the *basic affective tone* and rules out the possibility of a confound between our sublexical predictors and lexical values (emotional connotations of words used in the poems) holding responsible for the effects of *basic affective tone*.

Note that the model for arousal ratings differs a bit from the pattern described above, as a combination of the sublexical measure of valence and the corresponding control measure accounted for more variance than the sole sublexical measure of arousal. It is worth pointing out that in this model, when considering the bivariate or direct correlation between the dependent variable (i.e., the rating scores of arousal) and each single predictor, the sublexical measure of arousal has still the largest correlation with the rating scores when compared to other predictors. That is, the control measure of valence is added to the model due to its accounting for the residuals, acting as a suppressor variable, and not necessarily due to its own association with the rating scores. Note also that the control measure explains less variance than the measure based on salient phonological units and their valence also in this model.

SADNESS	Estimate	Step	p	R^2	simple corr	partial corr
Salient-SAV-Sigma-Arousal	.03	1	.0190	.09	.30*	.25
Control-SAV-Sigma-Valence	0	2			.07	
Control-SAV-Sigma-Arousal	0	3			.13	
Salient-SAV-Sigma-Valence	0	4			-.04	

SPITEFULNESS	Estimate	Step	p	R^2	simple corr	partial corr
Salient-SAV-Sigma-Arousal	.07	1	.0002	.22	.47***	.22
Control-SAV-Sigma-Arousal	0	2			.39**	
Control-SAV-Sigma-Valence	0	3			-.01	
Salient-SAV-Sigma-Valence	0	4			-.19	

FRIENDLINESS	Estimate	Step	p	R^2	simple corr	partial corr
Salient-SAV-Sigma-Arousal	-.05	1	.001	.17	-.42**	-.27
Control-SAV-Sigma-Valence	0	2			-.11	
Salient-SAV-Sigma-Valence	0	3			.16	
Control-SAV-Sigma-Arousal	0	4			-.19	

AROUSAL	Estimate	Step	p	R^2	simple corr	partial corr
Salient-SAV-Sigma-Arousal	.00	1	.49	.04	.22	.09
Control-SAV-Sigma-Valence	.08	2	.03	.06	.10	.29
Salient-SAV-Sigma-Valence	-.04	3	.03	.13	-.18	-.28
Control-SAV-Sigma-Arousal	0	4			.08	

VALENCE	Estimate	Step	p	R^2	simple corr	partial corr
Salient-SAV-Sigma-Arousal	.00	1	.0006	.19	-.44***	-.24
Control-SAV-Sigma-Valence	0	2			-.06	
Salient-SAV-Sigma-Valence	0	3			.18	
Control-SAV-Sigma-Arousal	0	4			-.26*	

*p<.05, **p<.01, ***p<.001

Salient-SAV-Sigma (-valence, -arousal): Sublexical measures of the *basic affective tone* (See Material & Method for details). *Control-SAV-Sigma (-valence, -arousal)*: Control measures calculated based on the whole phonemic inventory of a text to examine the potential of a circularity problem.

Table 11.2. Results of multiple-regression models for the prediction of ratings on five affective dimensions. Note that partial correlations are calculated based on the predictors that entered the stepwise regression model.

Moreover, when considering the estimated slopes of the models, or the respective correlation coefficients, it becomes obvious that significant sublexical predictors in each model invariably display the expected direction of correlation with the criterion of the ratings - as also previously revealed by the correlations between different rating scores themselves (see Table 11.1): valence and friendliness are negatively correlated with arousal, so the estimated slopes are in both cases negative, whereas perceived spitefulness and sadness increase with both arousal ratings and sublexical measures of *basic affective tone*. Similarly, the estimated slope for the sublexical measure of valence in the arousal model is negative and hence in accordance with the negative correlation between valence and arousal for the whole poems.

11.5 Discussion

In this article, we present a novel method for quantifying the *basic affective tone* of a text. The method is based on salient phonological units that might be used as foregrounded elements to produce or enhance affective and aesthetic effects during reading experience. For our analyses, we focused on the three contributing factors of communication–based on Jakobson's model of language function— that can also be applied to literary communication, i.e., author, reader and the text. By choosing the poems of Enzensberger's "verteidigung der wölfe" which the author himself assigned to three affective categories ("friendly", "sad", and "spiteful"), we were able to incorporate considerations about the author as a factor for statistical analyses. On the recipient side, we conducted an extensive rating study to assess readers' judgments on three affective, author-based dimensions "friendliness", "sadness" and "spitefulness", together with the dimensions of "valence" and "arousal" constituting the affective space of influential psychological emotion models (e.g., Barrett, 2006; Lang, 1995).

Results for the subjective ratings were –to a high extent—in accordance with the author's own categorization: the rating scores of the poems in the related category (e.g., rating scores of friendliness for the category of friendly poems) were always significantly higher than for the other two categories except for ratings on sadness in the sad condition differing significantly only from the friendly but not from the spiteful category, though displaying a respective tendency also in the latter case. This particular finding is better readable when considering the interrelation between the author-based affective categories and more general emotion dimensions: the sad and spiteful poems of “verteidigung der wölfe” might not represent totally distinct, but partially overlapping categories in a bi-dimensional affective space (poems from both categories received on average comparably negative valence ratings) – an assumption that receives further support from the rather disperse relation between sadness and arousal ratings across all poems. The respective partial difficulties to differentiate between the two categories (otherwise clearly distinctive concerning our results on spiteful and friendliness rating dimensions) become particularly understandable when considering the content of the poems: i.e., in the sad poems the mourning about the loss of nature and the ideal, and in the spiteful poems the rejection and ridicule of the very same, drawing here on the two categories of elegy and satire of Schiller’s poetic tripartite (Grimm, 1981).

Furthermore, since the two factors of author and reader are, naturally, related to the text, we focused especially on the latter exploring to which extend the use of phonological units determines the affective qualities of a literary text as perceived by the reader or created by the author. Based on a probabilistic model, developed in a previous study (Aryani et. al, 2013), we detected deviations between observed frequency and expected frequency of syllabic units in a given text, thereby extracting salient phonological units that appear significantly more often than expected. In a second step, emphasizing the “sound” properties of poetic language, we calculated *sublexical affective values* (*SAVs)* for each syllabic phonological unit; extending previous research on sound-meaning correlations (e.g., Heise, 1966) to a systematic analysis of a normative database comprising more than 5300 German words. Based on phonological salient units in a given text and their respective *SAVs* for valence and arousal, we developed a method to quantify the *basic affective tone* of a given text. This method includes a mean of inferential statistics, because it determines the extent to which the mean of *SAVs* of salient phonological units in a text exceeds a corresponding confidence interval that is derived from a corpus of 25 million words representing everyday language use (SUBTLEX, Brysbaert et al., 2011). The outcome of this method is a quantitative measure that allows making a statement–in terms of the level of significance–concerning the *basic affective tone* of a text based on affective valence and arousal at the level of salient syllabic phonological units. In doing so, our approach follows the notion that not poetic deviation by itself but the relation between deviating and standard distributed phonological units, between fore- and backgrounded elements, influences emotional reading experience (Jacobs, 2011; 2015a,b). This consideration of both, poetic and prosaic language as well as an explicit focus on quantitative differences concerning *SAVs* of salient, foregrounded units hence overcomes shortcomings of other studies that usually either used categorical affective values derived from rated word lists (Whissell, e.g., 2000, 2011), or, with an exclusive focus on poetic stimuli, do not take into account prosaic language at all.

The results of comparisons between the author-based affective categories revealed that differences between these categories –as evident in subjective ratings of emotional valence and arousal—could also be detected at the sublexical level. We could show, for instance, that our sublexical measure of the *basic affective tone* of the poems in regard to affective arousal is higher for the groups of spiteful and sad poems in comparison to the friendly poems. However, corresponding results for the sublexical measure of the *basic affective tone* in regard to affective valence did not differ significantly between the author-based affective categories. Therefore, the affective dimension of arousal seems to be a more influential in constituting the *basic affective tone* at the sublexical level than the one of valence – respectively its predictive power concerning the affective perception and evaluation of the whole text or poem. That is, with respect to our other results, an interesting outcome, to which we will return later.

As our approach to quantifying the *basic affective tone* of texts is based on the two dimensions of arousal and valence, we next divided the poems into two affective categories based on the median of their rating scores on I) arousal; contrasting a high- vs. a low-arousing category, and II) valence, contrasting a positive vs. a negative valence category. For each contrast (I&II), we could show that the respective *basic affective tone* differs significantly and in the expected direction between groups - whereas corresponding sublexical control measures did not display significant effects. Taken together, these findings clearly support the importance of phonological salience regarding the *basic affective tone* of poems.

The overall results of the regression analyses confirm these findings: our sublexical measures—and, again, not the control measures—significantly predict participants' ratings on each of our five emotion rating scales, suggesting that affective attributions of particular phonological structures influence the text's emotional perception by the reader reflected in the rating scores. The percentage of explained variance of the author-based affective dimensions amounts to 9.5% for the prediction of “sadness”, to 17.7% for “friendliness”, and to 22.5% for “spitefulness”.

Having a closer look upon our sublexical measures in the regression models, it became, again, obvious that the sublexical predictor of arousal (*Salient-SAV-Sigma-Arousal*) was almost always (in four of five cases) the best or the only significant predictor of the rating values. Our study, in line with a number of studies on the acoustic properties of emotional speech, hence provides support for an “acoustic arousal” dimension which claims that acoustic properties of speech provide vocal cues to the level of arousal, over and above valence (Bachorowski, 1999, Bänziger & Scherer, 2005, see also Sauter et al., 2010). Also, it has already been argued that vocal sounds primarily convey the arousal state of the sender (Bachorowski, 1999). This seems to be plausible when considering the psychological difference between valence and arousal. Arousal is related to a physiological state of being reactive to stimuli; it causes alertness and readiness, and involves more automatic and perceptual reactions, which in turn could be reflected in the vocal behavior of the sender and acoustic features of speech and written language. Valence, however, involves higher order, cognitive and evaluative processes (e.g., Briesemeister, Kuchinke, & Jacobs, 2014; Briesemeister, Kuchinke, Jacobs, & Braun, 2015; Jacobs et al., 2015; Recio et al., 2014) and might not be easy to detect at such a basal level as the phonological one. Based on our results, we therefore argue that at the sublexical level the affective dimension of arousal might be more suitable for measuring the *basic affective tone* than the affective dimension of valence.

The overall results of our study also support Jakobson's assumption that within poetry phonological "structures, particularly powerful at the subliminal level, can function without any assistance of logical judgment and patent knowledge both in the poet's creative work and in its perception" by the reader (Jakobson, 1970: 198). Our statistical operationalization provides strong evidence for the importance of 'sound' and supports the idea about a relation between sound and meaning as proclaimed by scholars and poets throughout history.

While the *basic affective tone* may, at the one hand, be understood as a stylistic device determining the "tone color" of a text at a holistic level, our statistical operationalization of *basic affective tone* offers, at the other hand, the possibility to capture a variety of more specific stylistic devices, or to investigate the effects of intonation patterns at higher levels of analysis (e.g., sentences, verses). Being based on over-proportionally used—and thus salient—units in a text, the here presented measure of *basic affective tone* may relate to all kinds of stylistic devices bearing on sound-patterning in the form of both simple repetitions (e.g., alliteration, assonance, and consonance) or at a higher level of design (e.g., chiasmus and envelope). As such sound patterns – when artfully employed – not only shape order in a text but help to emphasize its meaning, the *basic affective tone* may, therefore, in certain cases relate to secondary semantic effects (cf. Neuhäuser, 1991) achieved through stylistic devices.

Surely, the here presented study represents only a first step in investigating the *basic affective tone* of formed language. To be able to make statements about the actual influence of the *basic affective tone* on the reader, future research should consider the possibility to work in an experimental setting with texts that have been systematically manipulated at the sublexical level of language. A possible example for such manipulation could be the alternation of rhetorical features such as rhyme that lead to phonological recurrences (cf. Menninghaus et al., 2015).

Since our approach is based on the information in written texts read silently by participants, the question arises how the affective load of phonological units, as reflected in our statistical measure, the *basic affective tone*, can affect reader's perception. The answer lies in the process of automatic phonological and prosodic recoding of written words in silent reading (see Jacobs & Grainger, 1994, for an overview, and Jacobs et al., 1998, for a formal model) that should play an even more important role in poetry reception (Jacobs, 2011; Schrott & Jacobs, 2011). Research on visual word recognition in the last two decades has indeed provided accumulating behavioral, computational, and neuroimaging evidence that during silent reading phonological information is automatically generated from the printed word providing an early and major constraint for lexical access (e.g., Braun et al., 2009; Conrad et al., 2007, 2009, 2010; Ziegler & Jacobs, 1995; Ziegler et al., 2000; 2001).

Although the results of this study are so far restricted to Enzensberger's "verteidigung der wölfe", we want to note that our statistical quantification of the *basic affective tone* can by all means be applied to any other text form beyond the poetic genre. Especially the analyses of texts that are intended to elicit a certain affective impact in the reader, such as advertisements, political speeches or manifests, seem to be of special interest for future research. It is also worth mentioning that the here presented method is not limited to a specific language, rather it can easily be extended to any language for which comparable databases and corpora necessary to apply the method are accessible.

Finally, we suggest to use the here presented approach as a complementary method for sentiment analysis of written texts or transcribed speeches. So far, Computer-Assisted Text Analysis (CATA) usually link occurrence of certain words in a given text to the text's emotional content by using predefined keywords or word clusters. These methods, however, miss to consider the sound component of language. Our here presented approach represents a possibility for future sentiments analysis and opinion miming by expanding such approaches by an analysis of the *basic affective tone*.

III
General Discussion

Chapter 12

General Discussion and Outlook

The main goal of this dissertation was to investigate the role of affective iconicity in language processing and literature reception from empirical and theoretical perspectives. To gain a comprehensive insight into this broad topic, I employed a large variety of methods, including rating studies, decision tasks, fMRI, acoustic analysis, corpus analysis, and computational modeling. Moreover, I investigated this phenomenon in two different perceptual domains; visual (printed words) and auditory, and at two different processing levels; the single word, and the whole text (Table 12.1).

Study	Level	Modality	Method	Dep. Var.
1	Word	Visual & Auditory	Phon. & Acoustic Analysis	Rating
2	Word	Auditory	fMRI	BOLD signal
3	Word	Visual	Behavioral	Latency & Accuracy
4	Word	Auditory	fMRI	BOLD signal
5	Text	Visual	Corpus Analysis	Phon. Salience
6	Text	Visual	Comp. Modeling	Rating

Table 12.1. A summary of different textual levels (single word vs. text), and presentation modalities (auditory vs. visual), as well as different methods and dependent variables used in each of the empirical studies. Dep. Var. = dependent variable, phon. = phonological, comp. = computational

Study 1 made use of two rating studies for words and pseudowords with a large number of participants (N_{total} = 439) to provide novel measures for the affective sound of words, and in order to examine the effect of (implicit) sound on affective evaluation of meaning. Study 3 explored the role of affective sound in making semantic decisions about the affective meaning of words, using a novel two-alternative forced choice task on affective arousal (Arousal Decision Task). Study 2 and Study 4 made use of fMRI to investigate the neural correlates and substrates of the affective potential lying in the sound of words, and the interaction of sound and meaning, respectively. Study 5 developed a probabilistic model through a corpus analysis, thereby providing a text analysis tool, called EMOPHON, which automatically delivers information about sublexical phonological salience. Finally, Study 6 developed a quantitative measure for assessing the basic affective tone of texts, and provided a proof of concept by using a rating study, again, with a large number of participants (N = 252).

The presented studies account for three of four possible measurement classes proposed recently by Dixon and Bortolussi (2016), and displayed in Figure 12.1; indirect online (fMRI), indirect offline (response time), and direct offline (rating).

		Temporal Relationship			
		Online		**Offline**	
Inferential Relationship	**Indirect**	❑ eye movements, pupil dilation ❑ EEG, fMRI ❑ psychophysiological measures	✓	❑ recall ❑ judgment response time	✓
	Direct	❑ concurrent verbal protocols, ❑ probe responses ❑ text annotation	—	❑ questionnaires assessing reactions and experience ❑ interviews	✓

Figure 12.1. The present work covers three of four measurement classes for measuring literary experience, as suggested by Dixon and Bortolussi (2016).

A summary of the dissertation's main findings is illustrated in Figure 12.2, which I will discuss in detail in the next sections.

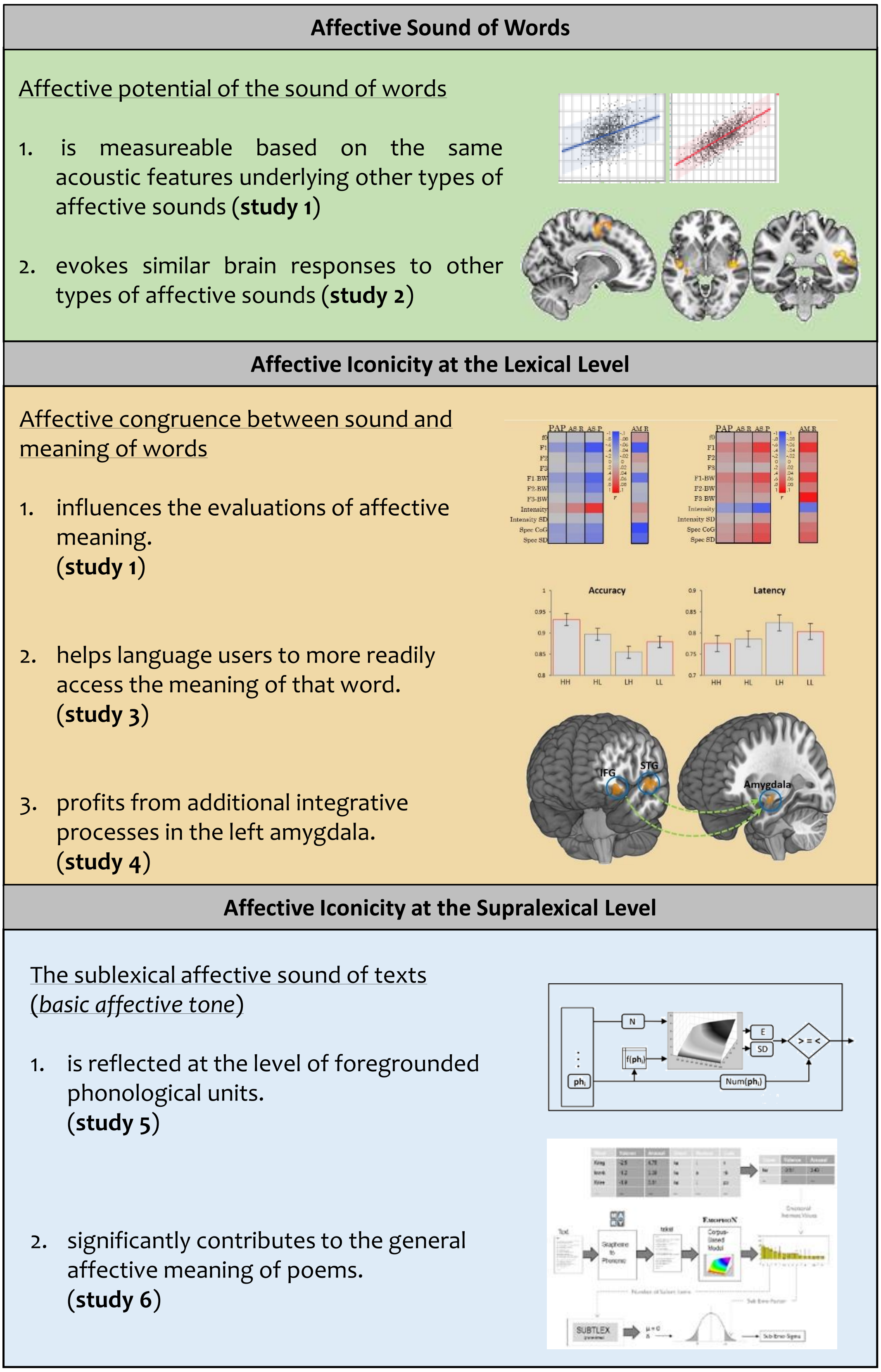


Figure 12.2. Results of the empirical studies are summarized in three main categories concerning the three initial main questions (Chapter 1) of this dissertation.

12.1 Affective Sound of Words

An important goal of this dissertation was to provide methods and tools for investigating and quantitatively measuring a phenomenon largely neglected in previous empirical research, namely the affective potential of the sound of words. This was investigated at both behavioral (Study 1) and neural processing level (Study 2), the results of which I will discuss in the following sections.

12.1.1 Psychological Reality of Affective Sound

In Study 1, results showed that language users can consistently evaluate affectivity in the sound of words. Words and pseudowords were rated for their affective valence and affective arousal. Importantly, the variance in rating values could be explained by means of a selected number of acoustic variables known for modulating nonverbal emotional vocalization and affective prosody. This indicates an inherent affective quality in the sound of words which goes beyond conventional links to linguistic concepts. Furthermore, results of this study laid the groundwork for a further investigation into this phenomenon by providing a reliable measure of the affective sound of words that we used in a number of follow-up studies. Results of analyses conducted in this study revealed a series of both acoustic variables and phonetic features that potentially underlie affectivity in the sound of words. The findings therefore provide insights into both the perception and the production of affectivity in the sound of words: the acoustic variables add to our understanding of how and based on which acoustic cues the affective impact of the sound is *perceived* and potentially evaluated, while the phonetic features and their underlying articulatory properties contribute to a better understanding of how such sounds are *produced*. I will now proceed to give a summary of these potential features related to these two levels of analyses and their likely role in the perception and the production of the affective sound of words.

<u>Acoustic Variables:</u> At the level of acoustic cues, *F1* and *F3* (the first and the third formants), *sound intensity*, and the *spectral center of gravity* (CoG) appeared to be the dominant features explaining the largest portion of variance in words' affective sound,

with F1 and CoG correlating positively, and sound intensity correlating negatively with the level of affective arousal (and positively with the level of affective valence) in the sound of words. These results align with previous investigations on acoustic cues in nonverbal emotional vocalization (Banse & Scherer, 1996; Frühholz & Grandjean, 2013; Juslin & Laukka, 2003; Sauter, Eisner, Calder, & Scott, 2010) and affective prosody (Belyk & Brown, 2013; Brück, Kreifelts, & Wildgruber, 2011; Juslin & Laukka, 2003), showing a high similarity of acoustic cues to affective judgments across different types of affective sounds (e.g., speech, music, and environmental sound) (Weninger, Eyben, Schuller, Mortillaro, & Scherer, 2013). In addition, this similarity supports those views proposing that the emotional content of sounds can be mapped out in terms of some basic acoustic cues (see Juslin & Laukka, 2003, for a review).

<u>Phonetic Features:</u> At the level of phonetic features, the results suggest that *short vowels*, as /ɪ/ in "sick" /sɪk/, compared to their long counterparts, as /iː/ in "seek" /siːk/, can make words sound more negative and arousing. Given the phylogenetic continuity of the vocal expression of emotion, which involves similar brain structures in both humans and other social mammals (Panksepp, 2000), it is insightful to note that the use of short calls in aggressive circumstances is an observable behavior in both (Morton, 1977).

Also, *voiceless consonants*, and particularly *plosive* ones (e.g., /p/, /t/, and /k/), can significantly contribute to making a word more negative and arousing. This type of phonemes are produced with an interruption and a following explosive release of the air stream which, in turn, can be associated with a higher level of arousal during its production.

Similarly, a highly arousing and negative affective impact was observed for *hissing sibilants* (e.g., /s/, /z/, and /ʃ/), which are strongly stressed consonants produced by a high-velocity jet of air against the teeth. This results in an arousing hissing sound which may account for the cross- and paralinguistic use of these sounds for attracting the attention of others (e.g., "psst!") as well as for their prominent deployment in literature as a stylistic device for cacophony.

These features can effectively be used for constructing words and pseudowords associated with specific affects (positive/negative, arousing/calming) or emotions (e.g., fear, disgust), which can have broad implications in various contexts from marketing (c.f. Schrott & Jacobs, 2011; Spence, 2012; Baxter & Lowrey, 2011) and advertising to art and literature.

12.1.2 Neurophysiological Reality of Affective Sound

Using the psychoacoustic model for measuring the affective sound of words provided in Study 1, my colleagues and I experimentally investigated its underlying neural correlates, and examined whether the affectivity in the sound of words can evoke affective responses in the brain (Study 2). The overall results of this study provided a comprehensive picture and the first evidence of the neural network processing of affectivity in the sound of words. In sum, this network consists of three major parts: i) superior temporal areas which are likely involved in the decoding of affective meaning from sound (cf. Ethofer et al., 2006; Wiethoff et al., 2008; Belin et al., 1998), ii) insular regions which are likely involved in the perception of sound (c.f. Frühholz, Trost, & Kotz, 2016; Mirz, Gjedde, Sødkilde-Jrgensen, & Pedersen, 2000; Trost, Ethofer, Zentner, & Vuilleumier, 2012), and iii) motor-related areas which are likely involved in emotional behavior in reaction to the sound (c.f. Löfberg, Julkunen, Pääkkönen, & Karhu, 2014; Zald & Pardo, 2002; Janata, Tomic, & Haberman, 2012). This response pattern fits perfectly to a proposed unifying neural network of affective sound processing (Frühholz et al., 2016), stating that all affective sounds induce brain activity consistently in a common core network. This is in opposition to the traditional view that proposes distinct neural systems for specific affective sound types.

12.1.3 Conclusion

Results of these two studies on the psychological and neurophysiological reality of the affective potential of the sound of words (Study 1 and Study 2) provide strong evidence that the specific sound profile of any word in a language can be attributed to a specific emotion as perceived by the listener. In Study 1, results showed that similar acoustic cues modulating affective prosody and nonverbal emotional vocalization shape the affective sound of words — as evident in the acoustic analysis of words and in the proposed psychoacoustic models based on pseudoword material. In Study 2, results showed that the affective sound of words evokes brain responses similar to those evoked by other types of affective sounds.

This observed similarity at both the behavioral and neural level provides empirical support for the Panksepp-Jakobson-Hypothesis as outlined in Chapter 3 (Jacobs, 2011; Jacobs 2015a). At the behavioral level, we even found acoustic cues that are similar between human subjects and other mammals in the context of emotional vocalization (e.g., short vowels vs. short calls). At the level of neural processing, in line with the Panksepp-Jakobson-Hypothesis and the general idea of 'neural reuse' (Anderson, 2010; Ponz et al., 2014; Ziegler et al., 2018), we observed the processing of affective sound in the same brain areas used for processing of other, more simple affective sounds.

However, despite the anticipated activation in the subcortical areas, as put forward by the Panksepp-Jakobson-Hypothesis, we also observed activations in, for instance, the auditory cortex, with its human-specific role in evaluative processes, such as in music-evoked fear (Koelsch et al., 2013), and the posterior insula, with its role in conscious feeling of affective sound; two functions that are processed in specialized brain areas not necessarily shared with other mammals. Therefore, it seems that the Panksepp-Jakobson-Hypothesis might need to be complemented with additional brain mechanisms responsible for aesthetic and affective processing of more complex stimuli. Based on these findings, I would like to suggest that the emotional and aesthetic experiences evoked by linguistic material and literature are not only processed in the brain structures that we share with other mammals, but also in brain areas that developed at earlier stages of human evolution for processing other, more basic types of stimuli. This view extends the brain mechanisms proposed by the Panksepp-Jakobson Hypothesis to those that are human-specific but originally evolved for accomplishing other tasks (Figure 12.3).

For instance, in the case of emotional vocalization, there are two systems underlying human vocal behavior. The first one, which can also be found in animal vocalization, is based in the limbic system (Freeman & Herron, 2007; MacLean, 1990; Panksepp, 2000; D. W. Ploog, 1992). The second one is cortically represented and began to emerge in primates but became qualitatively different in Homo-sapiens and is therefore unique to humans (Jürgens, 2003; Ploog, 2002). The proposed extension of the Panksepp-Jakopson-Hypothesis thus suggests that the perception of affective sound in words, for instance, makes use of both systems: a phylogenetic old system shared with other mammals, and a newer system that is cortically represented. Importantly, none of these systems evolved with the original purpose of evaluating such complex material (See Ziegler et al., 2018 for a recent study on the idea of neural reuse).

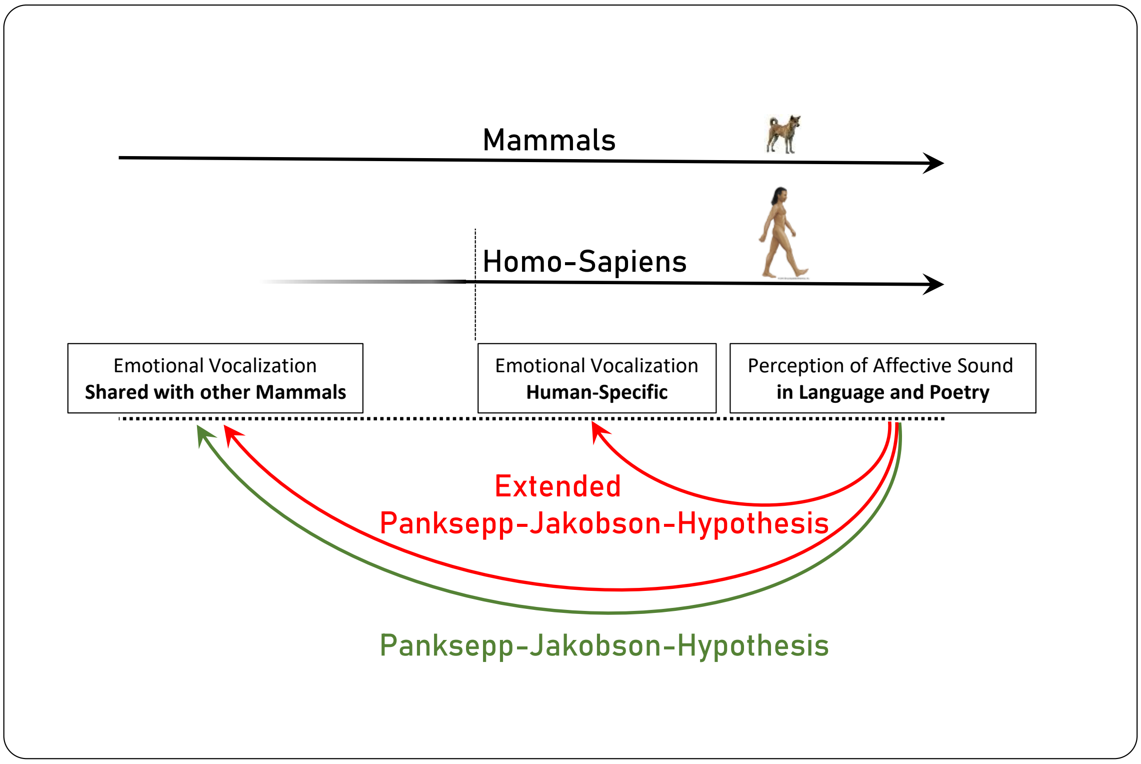


Figure 12.3. A proposed extension of the Panksepp-Jakobson-Hypothesis as exemplified for perception of affective sound in language and poetry

12.2 Affective Iconicity at the Lexical Level

A further goal of this dissertation was to examine the effect of sound on the lexico-semantic processing of words. As displayed in Figure 12.2 and Table 12.1, this was investigated for three experimental tasks (i.e., rating, semantic decision, and passive listening), two different modalities (i.e., visual and auditory), and at two prcessing levels (i.e., behavioral and neurophysiological), the results of which I will summarize in the following sections.

12.2.1 Affective Sound Influences Affective Evaluation of Meaning

When giving their affective judgments (valence and arousal) about the meaning of words, participants were implicitly influenced by the sound of words even when words were presented visually and read silently (as demonstrated in Study 1). Results of this study confirm my initial hypothesis (RQ 1) by showing that the implicit sound of words contributed to valence and arousal ratings so that a statistically significant portion of their variance could be accounted for by words' acoustic features after converting them into spoken form. Harsh-sounding words, for instance, were shown to make people feel more aroused so that they implicitly gave a higher arousal rating, even though they were instructed to only focus on the lexico-semantic aspect of words.

These results align with previous work on the interaction between the emotional significance of affective prosody and higher-order processes of meaning making—even though the attentional focus was not directly on the emotional cues of the sound (Brück et al., 2011; Frühholz, Trost, & Kotz, 2016; Grandjean et al., 2005; Kanske & Kotz, 2011; Kotz & Schwartze, 2010; Schirmer & Kotz, 2006). Similar to the results of this study, affective prosodic cues have even been shown to be engaged in silent reading by means of cross-sensory input from the visual cortex into the auditory cortex and affective regions of the brain (Berthier & Pulvermüller, 2011; Brück et al., 2014; Perrone-Bertolotti et al., 2012), and the phonologically recoded neuronal representation of the acoustic features corresponding to phonological word forms (Braun, Hutzler, Ziegler, Dambacher, & Jacobs, 2009; Breen, 2014; Mesgarani, Cheung, Johnson, & Chang, 2014; Ziegler & Jacobs, 1995). Therefore, in line with the role of affective prosody in meaning making, results of this study suggest that the process of affective lexical evaluation—as higher order cognition—is influenced by words' phonological form.

These results, together with the findings on phonetic features, can explain why some words which include specific sounds are more suitable for denoting specific concepts (e.g., swear words), and are therefore preferred by language. This may lead to a kind of "natural selection" that "makes some words more fit to survive" (Jespersen, 1922). The example of 'piss' vs. 'pee', as the title of Study 1 suggests, may perfectly illustrate this.

12.2.2 Affective Iconicity Facilitates Semantic Decision

A similarity between the sound and meaning of a word was hypothesized to help language users to more readily access the meaning of that word through direct form-meaning mapping. Inspired by the results of Study 1 on the influence of affective sound on the evaluation of affective meaning, we investigated the effect of sound on semantic decisions (Study 3) in two different groups of words: iconic vs. non-iconic words. These groups were defined as a function of the congruency between affective sound and affective meaning. In line with our initial hypothesis (RQ. 3), faster latencies and higher accuracy in responses were observed for iconic words when making a forced choice decision about the degree of arousal of presented words (high vs. low).

Results of this study extend those of the previous works on the facilitative effect of iconicity reported in sign language (Thompson et al., 2012; Vinson et al., 2015), and in ideophones (Dingemanse et al., 2016; Kita, 1997; Kwon & Round, 2014; Lockwood et al., 2016). By using quantitative measures for both the sound (Study 1) and meaning (Võ et al., 2009) of words, we tested the effect of iconicity for more ecologically valid stimuli and a larger number of 'ordinary' words in the lexicon.

This finding also supports the proposed role of multimodal convergence of emotions in making appropriate and faster decisions in emotional evaluation (see

Klasen, Chen, & Mathiak, 2012; Schirmer & Adolphs, 2017, for recent reviews). Previous investigations in this field, which have primarily used emotional faces and voice, have shown that the presentation of congruent bimodal emotional cues (both face and voice) yields faster and more accurate emotional judgments than unimodal presentations (e.g., only face; Calvert, 2001; Schirmer & Adolphs, 2017; Schröger & Widmann, 1998). In a similar fashion, words possessing congruent affective information from two different sources (i.e., sound and meaning) were, in study 3, categorized more quickly and more accurately.

An important finding of this study is the possible affective and aesthetic value of iconicity in language and, in particular, in literary reading (Jacobs, 2015a; Schrott & Jacobs, 2011). As discussed in Study 3, the facilitated lexical processing of iconic words can be linked to previous findings on the notion of processing fluency, which have demonstrated that a greater ease of processing leads to a higher level of aesthetic pleasure (Bohrn, Altmann, Lubrich, Menninghaus, & Jacobs, 2013; Menninghaus et al., 2014; Reber et al., 2004; Song & Schwarz, 2009). This may provide a novel explanation for the aesthetic impact and common use of stylistic devices such as phonaesthetics and iconicity in poetry (Jacobs, 2015a; Schrott & Jacobs, 2011; Ullrich et al., 2017; Whissell, 1999). I will discuss this potential aesthetic function of iconicity and its relevance to the model of literary reading in more detail in a later section of this dissertation.

12.2.3 Affective Iconicity Profits from Additional Neural Mechanisms

As summarized in Chapter 4, a wealth of previous studies on iconicity have suggested that iconic words facilitate semantic decisions through the activation of additional links between the sound of a word and modality-specific experiences (i.e., sensory, motor, affective), as well as through the integration of information from different modalities, which may provide an opportunity for the stronger embodiment of iconic signs. In line with this hypothesis (RQ 4), results of Study 4 showed that iconic words—organized very similarly to how they were organized in Study 3—resulted in an enhanced hemodynamic response in the left amygdala, an area of the brain known for its role in supramodal emotion integration, and as a general ‘convergence zone’ (Park et al., 2008; Schiller et al., 2009; Kreifelts et al., 2010; Klasen et al., 2011; Müller et al., 2011). Furthermore, an Psycho-Physio-Physiological-Interaction (PPPI) analysis with the left STG and the left IFG as seeds revealed a significant functional connectivity with the left amygdala in the iconic condition.

Showing a greater engagement of affective brain regions for (affective) iconic words, our findings can advance understanding of affective and aesthetic processes of literary reading (Jacobs, 2015a; Schrott & Jacobs, 2011). In line with its role in multimodal emotion integration, it has been proposed that the left amygdala responds to metaphoric language, valence congruity, figurativeness, and harmony (Jacobs, Hofmann, et al., 2016). Empirical support for this view comes from studies showing enhanced activation of the left amygdala for metaphors (Citron & Goldberg, 2014) and metaphorical Noun-Noun-Compounds (Forgács et al., 2012) when compared to their literal counterparts. Also, results of a meta-analysis of 23 neuroimaging studies showed a left amygdala activation in response to a variety of figurative statements, and in particular metaphors (Bohrn, Altmann, & Jacobs, 2012a). The meaning of metaphors is generally based on considerations of similarity between different aspects of target and source, and this is what iconicity in language is about. Lakoff and Turner (Lakoff & Turner, 1989) defined iconicity as a “metaphorical image-mapping in which the structure of the meaning is understood in terms of the structure of the form of the language presenting the meaning”. Such image-mapping, according to them, is enabled by image-schemas which are formed from our embodied experience. This view emphasized the role of the left amygdala as a central hub critical for regulating the flow and the integration of information from different experiences.

12.2.4 Conclusion

These results challenge the established notion of arbitrariness put forward in the Saussurean model of language, and they support the more textured models proposed by Peirce, Bühler, and Jakobson. The results thus provide empirical evidence for an insightful prediction that Jakobson (1937) made almost 80 years ago: “the intimacy of connection between the sounds and the meaning of a word gives rise to the desire of speakers to add an internal relation to the external relation, resemblance to contiguity, to complement the signified by a rudimentary image".

By investigating the effect of iconicity on different processing levels, i.e., from an entirely explicit task of rating, through to semantic decision, and finally the implicit passive listening, my colleagues and I could provide firm evidence for iconicity as a general feature of language (cf. Perniss et al., 2010). As I outlined in the introduction, unlike pioneering works on the facilitative effect of iconicity in sign language (Thompson et al., 2012; Vinson et al., 2015) which laid the foundation for the theoretical framework of such investigation, related research on spoken language has generally suffered from a number of limitations and methodological shortcomings that have already been addressed in this dissertation project.

The first limitation was a predominant use of pseudowords instead of real words. By using real words in all of the relevant studies (Study1, Study 3, and Study 4), it is now possible to directly apply the results of these studies to the models of natural language processing, thereby improving our understanding of the effect of sound on the process of meaning making.

Furthermore, the results of behavioral studies and acoustic analyses, together with the insights form the neuroimaging studies, helped overcome a major limitation of previous work regarding the nature of the relation between sound and meaning, namely the question of iconicity vs. systematicity (Dingemanse et al., 2015). A sound-meaning mapping is considered iconic when both sound and meaning independently refer to a similar specific (sensory, motor, or affective) domain. In all of our studies, we used two different measures for assessing the sound and meaning of words based on their affective arousal. At the meaning level, our measure for the lexical arousal (Võ et al., 2009) was cross-validated in various empirical studies regarding experiential, behavioral, and neurobiological levels of analysis (Jacobs et al., 2015). At the sound level, the measure of sublexical arousal used in our studies has been shown to i) have an inherent affective quality based on acoustic features that are known to modulate nonverbal emotional communication (Study 1) and ii) can evoke affective brain responses similar to other types of affective sounds (Study 2).

Consequently, it is reasonable to conclude that our findings on the effect of sound-meaning mapping are related to iconic mappings of words rather than statistical regularities in the lexicon.

Finally, investigating the role of the affect and affective meaning of words, we moved beyond the narrow focus on limited semantic concepts. This enabled us to test sound-meaning correspondences across a large number of words which were more representative of the lexicon. In addition to a wider semantic space, the quantitative operationalization of meaning and sound helped us to test the effect of iconicity for a more ecologically valid material, namely the 'ordinary' words in the lexicon.

12.2.5 A Neurocognitive Model of Affective Iconicity

Results of Studies 1-4 are integrated into a model of visual word processing for affective iconicity as presented in Figure 12.4. Classic models of visual word processing suggest that affective evaluation of words is based on the semantic content only, which is linked to the phonological and orthographical representation of words in semantic memory. However, the reported results and the overall findings of this dissertation suggest an additional processing route beside the associative links between phonology and semantics. Results of Study 1 and Study 2 indicate that the sound of words has its own affective impact on the reader/listener. Specific phonological cues (e.g., voicing, vowel length, hissing, etc.) and acoustic features (e.g., intensity, CoG, formants, etc.) are processed and evaluated in specific brain regions, with the STG playing the most significant role for the evaluation of the (implicit) sound per se. As suggested by the results of Study 1 and Study 3, the affective impact of the sound of words can be integrated with higher-order cognitive processes related to semantic content of words and thereby influence the final affective evaluation. Such an integration can take place at both processing levels of i) affective rating, resulting in an enhanced or diminished rating value (Study 1), and ii) semantic decisions, resulting in faster or slower response times as well as higher or lower accuracy in making semantic decisions as a function of the congruence or incongruence of the sound and meaning (Study 3). That is, the affective potential of the sound can both positively and negatively interact with the initial affective meaning derived from the semantic content of words. The results of both the categorical comparison and functional connectivity from Study 4 revealed the locus of the effects: The integration of these two factors—affective sound and the initial affective meaning—will take place in the left amygdala.

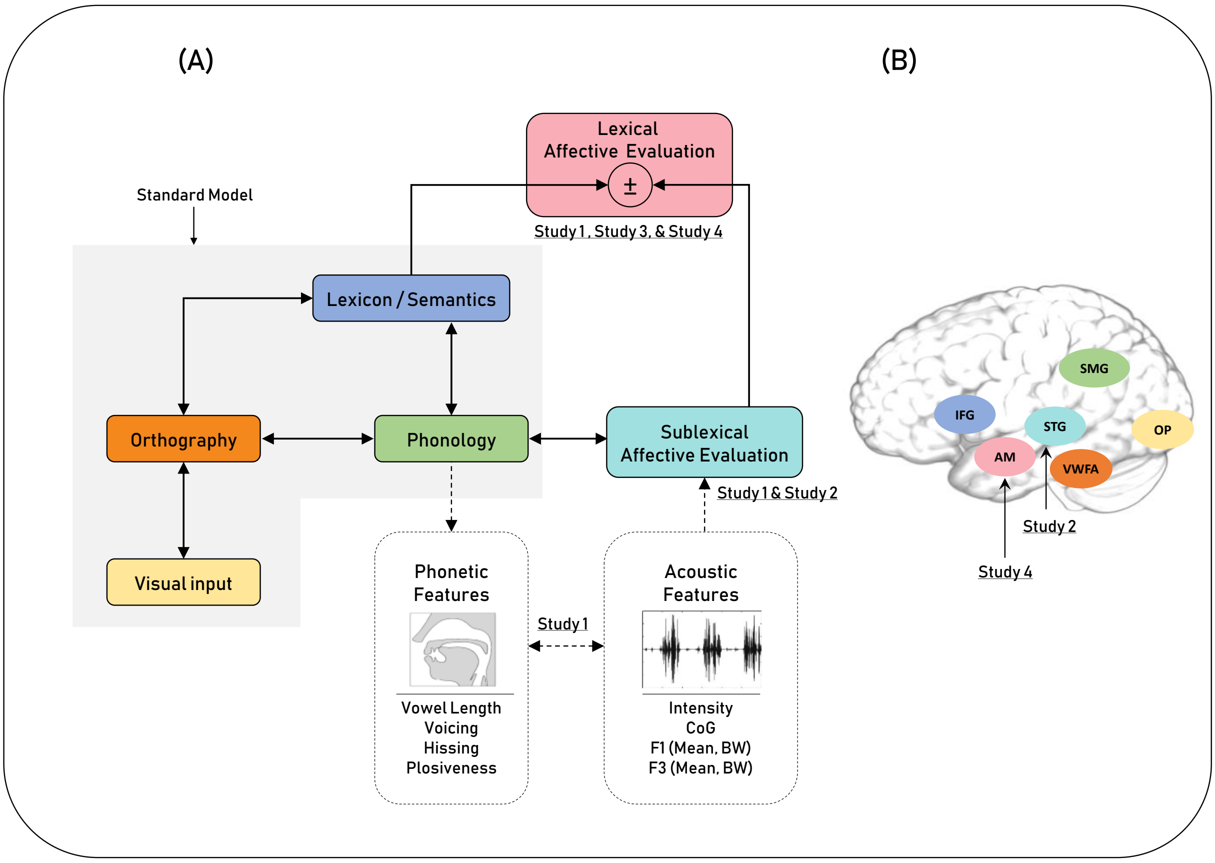


Figure 12.4. A neurocognitive model of affective iconicity. **A)** The extension of a standard model of visual word processing to the effect of affective iconicity. **B)** The most critical brain regions associated with different representations. The color of the ovals denotes the theoretical representations in the model. Abbreviations: IFG, inferior frontal gyrus; SMG, supramarginal gyrus; STG, superior temporal gyrus; OP, occipital pole; AM, amygdala; VWFA, visual word from area.

12.2.6 Valence and Arousal

A crucial finding across all of the studies in this dissertation project was a consistent dominant role of arousal, compared to valence, in iconic mappings and in shaping the affective sound of words. In Study 1, for instance, there was a much larger degree of agreement among raters (as measured through ICCs) for arousal than valence when evaluating affective sound of both words and pseudowords. This was, in turn, reflected in a significantly larger amount of variance in arousal ratings than valence ratings that could be accounted for by acoustic features. Previous cross-language investigations in sound-meaning correspondences conducted by my colleagues and I revealed a similar role for arousal—compared to valence—across all different languages in question, i.e., German, English, and Spanish (Conrad, Aryani, & Jacobs, unpublished work). Also, at the level of a whole text, the sublexical measure of basic affective tone for arousal—compared to that for valence—was (in four of five cases) the best or the only significant predictor of the rating values.

At the sublexical level, this finding supports a number of studies on the acoustic properties of emotional speech and affective sounds. These studies claim that acoustic properties of speech provide vocal cues to the level of arousal, over and above valence (Bachorowski, 1999; Bänziger & Scherer, 2005, Sauter et al., 2010). According to this view, vocal sounds primarily convey the arousal state of the sender (Bachorowski, 1999). Current research on vocal communication supports this notion. A handful of encoding studies show that actors vocally enacting a relatively small number of basic emotions produce differentiated patterns of vocal parameters for different emotions, among which arousal-related parameters play a significantly more important role (Bänziger, Hosoya, & Scherer, 2015). In other words, arousal differences in vocal emotion expressions are well captured by acoustic variables and voice ratings, and play a powerful role in the listeners' inference of the affectivity of sound. In a recent study using a comprehensive path model of vocal emotion communication, Bänziger et al. (2015) modeled data sets on emotional expression and recognition from two different cultures and languages. Results of their Lens model equations, hierarchical regression, and multivariate path analysis, reflect the strong evidence from past work in this field suggesting that the "voice is the privileged modality for the expression and communication of arousal and activation, whereas the face is vastly superior with respect to valence" (Scherer, Clark-Polner, & Mortillaro, 2011).

These results seem plausible given that more affective arousal is likely to be translated into increased vocal effort and faster speech (Banse & Scherer, 1996; Ethofer, Van De Ville, Scherer, & Vuilleumier, 2009; Juslin & Laukka, 2003; Scherer, 2003) which in turn affects many vocal cues. When considering the psychological difference between valence and arousal, the same conclusion on the priority of arousal at the sound level can be made: Arousal is related to a physiological state of being reactive to stimuli; it causes alertness and readiness and involves more automatic and perceptual reactions. This can be easily reflected in the vocal behavior of the sender and the acoustic features of speech. Valence, however, involves higher order, cognitive and evaluative processes (Briesemeister et al., 2014, 2015; Citron, 2012; Jacobs, 2015a; Jacobs, Hofmann, et al., 2016; Kuhlmann et al., 2016; Recio, Conrad, Hansen, & Jacobs, 2014) and might not be easy to detect at such a basic level as the phonological. This view aligns with Panksepp's hierarchical theory of emotions (Panksepp, 1998), which states that valence relies on phylogenetically younger brain structures and is computed at the so-called tertiary (i.e., neocortical) level of affective processing, thus reflecting higher-order categorization, reorganization, and appraisal processes. This is in line with previous work showing that the first emotional appraisal of a stimulus is related to arousal and not to valence, and that arousal is the primary factor producing emotional interference in information processing tasks such as the emotional Stroop or attentional blink paradigms (Anderson, 2005; Dresler et al., 2009; Schimmack, 2005). This can also explain the reason of a common failure of earlier attempts to identify a set of vocal features that reliably differentiate between the levels of valence (Bachorowski, 1999; Juslin & Laukka, 2003; Sauter et al., 2010).

12.3 Affective Iconicity at the Supralexical Level

Building upon the idea of foregrounding, we attempted to measure the sublexical affective sound of texts, termed "basic affective tone", by focusing on the foregrounded phonological units which we extracted from texts in a first step (Study 5), and then used—combined with their affective values—to deliver a statistical measure for the concept in question.

12.3.1 How to Extract Foregrounded Units?

Results of Study 5 were integrated into a computer-linguistic tool called EMOPHON, which automatically transforms texts into phonetic transcriptions, the absolute frequencies of which are documented and compared to expectation values derived from a large scale database of 25 Million German words (i.e., SUBLTLEX-DE; Brysbaert et al., 2011) representing everyday language use.

For a given text, EMOPHON provides the information of whether and to which degree any specific phonological unit occurs more or less often than would be expected. The expected values are an object of a statistical inference process assigning levels of significance for deviations based on a probabilistic model. EMOPHON therefore lays the foundation for all kinds of empirical research relying on the assumption that what is generally understood as higher level intentional message or the "basic affective tone" of a text may already be observable at the sublexical level. Once a phonological unit is identified as possessing a salient status (i.e., over-proportionally used), it is reasonable to assume that the specific unit might be used as an element of foregrounding – especially in the artistic domain of language use (e.g., poetry and literature) where the connection between the formal aspect of language and higher meaning is of great importance (Jacobs, 2015a; Jakobson, 1965; Jakobson & Waugh, 1979; Schrott & Jacobs, 2011).

Results of the comparison of phonological foregrounded units between poems and prosaic texts confirmed the hypothesis that the deviation of poetic language from norms characterizing ordinary language use is observable at the phonological level and can be measured by the numbers of foregrounded phonological units. The use of rhyme and other sound effects such as alliteration, consonance, and assonance can obviously affect the sound structure of the text, and, consequently, the iconic sound-meaning mappings in the form of foregrounded elements, which can be measured by the use of this tool. I will discuss this in more detail at a later point in this dissertation.

12.3.2 How Are Supralexical and Sublexical Levels Connected to each Other?

Focusing on the foregrounded phonological units in a text substantially reduces the complexity of analysis at the supralexical level by taking into consideration a limited number of phonemes identified as foregrounded rather than a larger number of textual features. This step was complemented in Study 6 by assigning affective values to these foregrounded features, providing statistical measures for assessing the "basic affective tone" of a given text. The method for calculating affective values for phonological units was very similar to that used for the calculation of *Phonological Affective Values* in Study 1. This was inspired by a classic study by Heise (1966), in which he assumed affective meaning of language units "at the macrolevel" co-varies with phonological structure "at the microlevel".

Beside the use of foregrounded elements and their iconic associations, an important advance presented in this method was to set a frame for interpretation of specific results independent from the results of other texts. That is, the provided statistical measure for each text is tested against a null model depending on the number of salient units found in the text. The null model was calculated by randomly pulling numerous chunks (1 million) of phonological units from the same corpus used in EMOPHON (i.e., SUBTLEX-DE). This enabled us to predict the affective load of phonological structures in a single poem without the necessity of further comparisons, overcoming one of the common limitations shared by almost all previous investigations.

Based on the NCPM, we focused on the three contributing factors of communication, i.e., the author, the reader and the text, and applied the method on

Enzensberger's "verteidigung der wölfe". We could therefore compare the obtained values for the basic affective tone between the author-based affective categories (i.e., spiteful, sad, and friendly). This revealed a similar affective categorization at the sublexical level to that defined by the author. Similarly, results of multiple regression models showed that a considerable part of the variance (9.5%-20%) of readers' ratings on different scales of emotion could be accounted for by the use of provided sublexical measures for basic affective tone.

Crucially, in both categorical comparisons and regression analyses, results showed that the respective measures of basic affective tone, which were based on foregrounded phonological units, differed significantly between groups, whereas corresponding sublexical control measures, which were based on the entire phonological material in the text, did not display significant effects. The overall results of the regression analyses provided further support for these findings: Sublexical measures based on foregrounded units—and, again, not the control measures—significantly predicted participants' ratings on each of our five rating scales of emotion.

Taken together, these findings clearly support the importance of phonological salience regarding the basic affective tone of poems, suggesting that affective attributions of *particular* phonological structures influence the text's affective perception by the reader.

12.3.3 Conclusion

Results of these two studies (Study 5 and Study 6) clearly support the initial hypotheses that iconicity is a potential indicator of the affective qualities of a literary text. The presented method provided some improvements that address a number of limitations of previous studies into the effect of iconicity in the literary genre of poetry (Albers, 2008; Auracher et al., 2010; Jakobson & Waugh, 1979/2002; Fónagy, 1961; Schrott & Jacobs, 2011; Tsur, 1992b; Whissell 1999, 2000, 2011).

As I previously pointed out, it should be emphasized that the presented method for measuring the basic affective tone is capable of capturing a variety of more specific stylistic devices. Although the basic affective tone may be understood as a stylistic device determining the “tone color” of a text at a holistic level, the statistical operationalization of basic affective tone offers the possibility of capturing different devices or investigating the effects of intonation patterns at higher levels of analysis (e.g., sentences, verses). This method can thus relate to all kinds of stylistic devices based on sound-patterning in the form of both simple repetitions (e.g., alliteration, assonance, and consonance) or at a higher level of design (e.g., chiasmus and envelope). As such, sound patterns—when artfully employed—can not only shape order in a text (Jakobson, 1965) but can also help to emphasize its meaning. In certain cases, the basic affective tone may therefore relate to secondary semantic effects (cf. Neuhäuser, 1991, Jacobs 2015a, 2015b) achieved through stylistic devices.

12.3.4 A More Dynamic Model of Literary Reading

The resulting effects of affective iconicity suggest a revision of the NCPM with a more detailed explanation of the dynamic nature of affective experience during reading literature and poetry.

An advanced property of the NCPM is that it makes predictions at both behavioral and neural level based on the elementary factors of the text, the context, and the reader. However, the structure of the model does not take into consideration the dynamic, temporal processes that occur in the model itself, i.e., the internal states of the model that are achieved during reading. More specifically, the model seems to lack feedback loops that can potentially change internal states of the model (e.g., emotional experience) based on its current outcome (e.g., reading behavior). For instance, in the case of iconicity, results of Study 3 and Study 4 highlight a potential association between iconicity and a feeling of aesthetic pleasure, evidenced by quicker responses in the semantic decision task, and the enhanced activation of the left amygdala in response to iconic words. Therefore, it is plausible to assume that the use of this technique at the sublexical level can influence the feelings evoked by the general theme of the text at the supralexical level. Imagine, for instance, a poem with a fearful theme and a general negative atmosphere. The use of iconicity at different textual levels in such a text can lead to more fluent reading, and this, in turn, can evoke positive feelings of aesthetic pleasure in the reader.

The prominent poem “The Raven” by Edgar Allan Poe, which possesses a great deal of iconicity and other sound devices, may serve as an excellent example for this interactive relation between different feelings evoked by different textual layers. Set within a sad and fearful atmosphere, the poem tells of a talking raven’s mysterious visit to a desperate lover, tracing the man’s slow descent into madness. Following the Fiction-Feeling Hypothesis, the general theme and the story of the desperate lover would be expected to provoke readers’ empathy with him and, consequently, engage the affective empathy network of the brain (Jacobs and Schrott, 2015; Jacobs, 2015a; Jacobs and Lüdtcke, 2017). That is, the descriptions of the protagonist’s pain and the dominant fearful atmosphere would cause increasing involvement of the core structure of pain and fear. Suddenly, the sad narrator hears a sound:

…*While I nodded, nearly napping, suddenly there came a tapping,*
As of some one gently rapping, rapping at my chamber door.
“’Tis some visitor,” I muttered, “tapping at my chamber door—
Only this and nothing more.”

Further to the structural similarities between the physical events (e.g., someone taps on the door several times) and syntactical features (the repetition of the word ‘tapping’), both the words *rapping* and *tapping* are iconic in that they resemble the sound of knuckles against a wooden door. The ease of processing of these iconic words and iconic structures would evoke a feeling of aesthetic pleasure, lust and play, which, in turn, would lead to the enjoyment of reading (Figure 12.5). The reader may want to read this stanza again in order to hear “the echo to the sense” (Pope, 1711) which is created by the sound-meaning interplay caused by iconicity. And this, in turn, may draw the reader out of fluent reading modus. The reader of these lines would therefore experience mixed, fluctuating feelings.

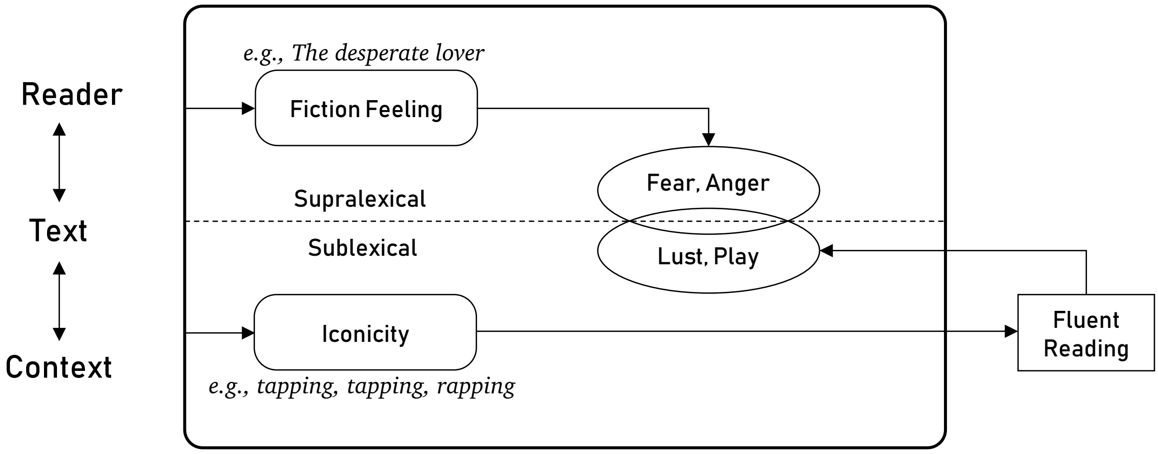


Figure 12.5. The interaction of supralexical and sublexical processes leading to different emotional experiences as exemplified by Poe's "The Raven".

This example shows how different textual layers (e.g., sublexical and supralexical) can influence reading behavior in a parallel fashion, and yet in opposite directions. Furthermore, it shows how the reading behavior can influence the experienced emotions during reading.

Taken together, I would like to propose a feedback loop between the behavioral level of reading and the emotional experience, as this is currently lacking in the NCPM. The model needs to be revised to take into account that different types of reading behavior (e.g., fluent vs. dysfluent) can potentially evoke specific emotions in the reader. I have illustrated this possible connection in the model by adding a general link between all different types of readings and related emotions in Figure 12.6 (The black arrow). More specifically, I will suggest a feedback loop between fluent reading and perceived emotions of lust and play (the red arrow) based on the findings of this dissertation project. A hypothetical link between dysfleunt reading and negative feelings (e.g., anger, fear) could also be speculated (the blue dashed arrow), but currently lacks empirical support.

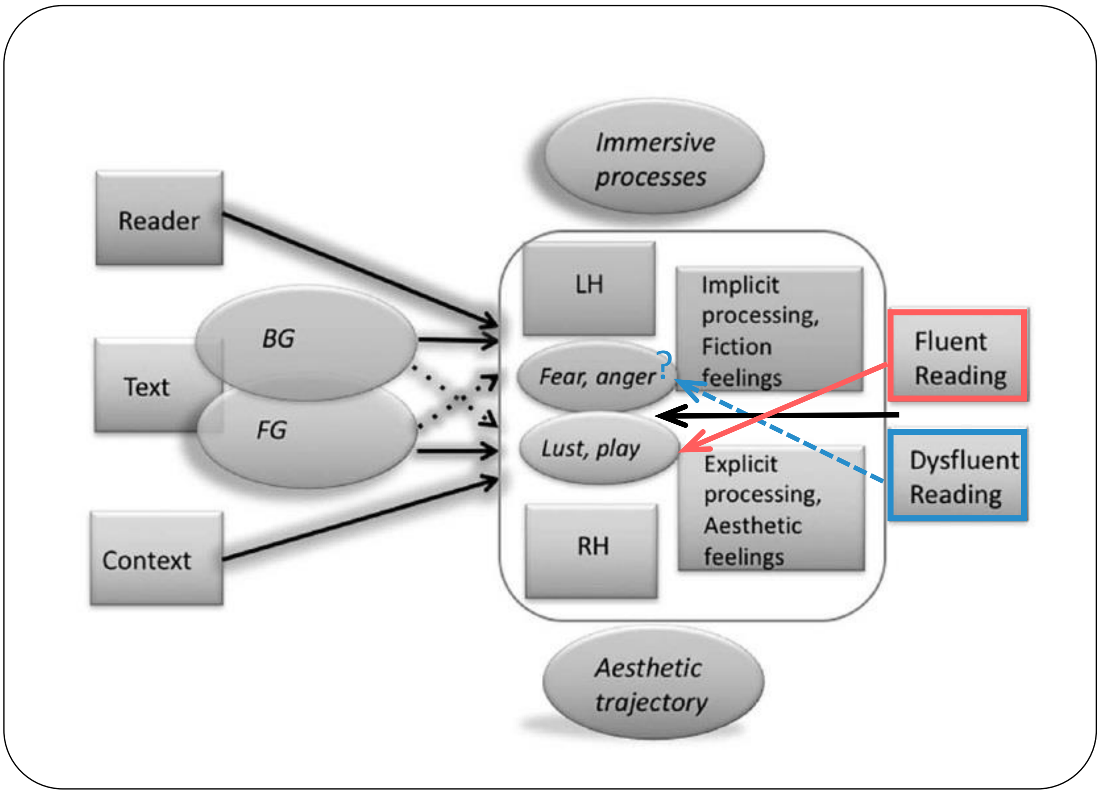


Figure 12.6. An updated version of the NCPM that accounts for the feedback loop from the reading behavior back to the emotional experience during reading.

12.4 Limitations and Outlook

I will conclude this dissertation with a discussion of some limitations and open questions that can extend the findings and conclusions of the present work and, presumably, challenge some of them. I shall then give an overview about some perspectives and suggestions that can be used in future research to overcome these limitations.

A general concern about the limitations of empirical studies in this work relates to a prevalence of computational methods throughout recent years. That is, machine learning methods have become increasingly popular, particularly for applications in complex data, including cross-validated out-of-sample predictions, which use different types of classification algorithms and methods (e.g., Jordan & Mitchell, 2015; Vogt, 2018). These methods are of a direct relevance to the present work, as they can improve the methods used in the reported empirical studies at different levels of sublexical, lexical and supralexical analysis, as well as the neuroimaging data.

12.4.1 What makes the sound of words affective?

With Study 1, a measure for assessing the affective sound of words was provided. This laid the groundwork for further investigations of sound-meaning correspondences in the present work. The best and most promising method was based on the ratings of pseudowords and linear regression models that could predict up to 58% of the variance in these rating values. This method encounters a number of limitations that can be improved in future research, as described in the following suggestions.

Firstly, the use of static acoustic features precludes the relevant information available in dynamic features. Here, the use of more dynamic sound variables such as spectral flex and sound entropy would increase the accuracy of acoustic models predicting ratings of affective sound. A more sophisticated approach might use the matrix of the entire spectrogram to quantitatively represent the sound envelope. Here, the representational similarity analysis (RSA) approach (Cichy, Pantazis, & Oliva, 2014; Kriegeskorte, Mur, & Bandettini, 2008) can be adopted to evaluate the representational content of the acoustic signal by correlating the differences between values in amplitude and frequency of two items with the differences resulting from the ratings values of these items.

Secondly, the sole focus on acoustic features leads to a neglect of features related to the production of the sound, i.e., phonetic features. A phonetic approach can complement the use of acoustic features. Here, for instance, words can be described as a concatenation of vectors of phonetic features corresponding to each phoneme in the word. As outlined in Study 1, a practical approach for defining such phonological cues could be based on the proportion of consonants with particular manner and place features, and the average height and position of vowels (as provided in Monaghan et al., 2007). An advantage of this method would be the simple classification of the phonological construction of a word and its contribution to the specific affective state.

Finally, the insights of sonority theory (Clements, 1990; Stenneken et al., 2005) can provide additional improvements for the measurement of the affective sound of words. Each item can be assigned a sonority score which can systematically contribute to affective (and aesthetic) ratings. In an initial investigation based on the mean and the standard deviation of sonority scores of different phonemes in an item (from 0 for voiceless plosive/affricate, through 4 for nasals, up to 7 for vowels), I could account for almost 15% of variance in affective ratings of arousal for the same pseudowords as in Study 1. Recently, this method has been successfully used in two different studies concerning ratings of the aptness of metaphors and the beauty of words (Jacobs & Kinder, 2018; Jacobs, 2017).

Specifically, the results of the study referenced above (Jacobs, 2017) can effectively guide future research on this topic. It is the first study that combines eight different QNA-driven features (including the novel metric of "aesthetic potential") with machine learning algorithms, and it shows that while none of the eight features on its own accounts for much variance in the data they fit very well together when predicting word beauty. The latter insight should motivate future investigations toward the combined use of different cues as suggested above, i.e., a combination of acoustic, phonetic, and sonority-based features together.

12.4.2 How sound contributes to affective meaning?

An important method used throughout this work at both lexical (Study 1) and supralexical levels (Study 6) was the calculation of a *Phonological Affective Value* for each of the phonemes or phoneme clusters. This calculation was, however, based on two simplifications that can be modeled more precisely in future research.

Firstly, when modeling the hypothesis that the affective potential of each phoneme can contribute to the rating of affective meaning of the entire word, all phonemes were weighted similarly regardless of their position in the word. A more precise model should take this factor into consideration, particularly because the beginnings of words have been shown to play a crucial role in visual word recognition (Jonathan Grainger & Jacobs, 1999; Jacobs, Grainger, & Ferrand, 1995) and in providing cues to semantic category (Durieux & Gillis, 2001; Kelly, 1992).

Similarly, the contribution of single phonemes to affective ratings was modeled in a simple additive way. It is, however, possible that this contribution would change depending on the semantic content and the affective connotation of the word. Thus, the effect of phonemes and the effect of semantic content can have an interactive effect on ratings of affective meaning. This was not considered in the presented model. Since the number of phonemes and semantic features can lead to a large number of different interactions, applying more sophisticated methods, such as machine-learning-based regressors (e.g., Jacobs & Kinder, 2018; Jacobs et al., 2017) can help to overcome the problem of high dimensionality, which results in more complete and accurate models.

As suggested above, a similar method based on the phonological cues (e.g, proportion of plosives, fricatives, etc.) can be used for the lexical level of analysis, complementing the acoustic approach employed in this work. In an attempt to predict the Phonological Affective Potential (PAP) by means of such phonological cues, I focused on the same words used in Study 1 and computed a list of 24 phonological cues for each of the words. An astonishingly large portion of variance, i.e., 74%, in the PAPs of arousal could be accounted for by 21 variables that successfully entered the stepwise regression analysis (compared to 11 acoustic variables accounting for 28% of variance in Study 1). These results should further encourage future research to combine different methods and variables to provide more sophisticated models of the role of sublexical features in the process of meaning making.

12.4.3 How neurons respond to affective iconicity?

In this work, the neural correlates and substrates underlying the affective sound of words and its interaction with affective meaning were addressed in two fMRI studies. However, due to the exploratory nature of the studies and the use of correlational methods, as well as the complexity of the underlying processes, a number of questions remain unanswered.

For instance, the temporal or causal relationship between respective neural networks and how precisely they interact with one another is still largely unknown. Although the present results have provided a first insight into the interactive relationship between iconicity and related brain regions by conducting functional connectivity analysis (i.e., PPI), employing advanced methods such as effective connectivity would help identify causal inference, or the use of combined EEG-fMRI, or MEG would both provide better temporal resolution. Furthermore, electrocorticography (ECoG) would deliver high resolution at both the spatial and temporal level (e.g., Mesgarani et al., 2014; Ponz et al., 2014), and would potentially help to overcome a number of limitations of the current data.

Also, more advanced analysis methods such as representational similarity analysis (Kriegeskorte et al., 2008) could be used in order to uncover underlying properties representing the affective sound of words. For instance, one proposal would be to investigate whether language users respond to the affective sound of words in a similar way to other affective stimuli, such as environmental sounds. For this, each word and each environmental sound could be modeled as separate events and representational dissimilarity matrices (RDMs) could be computed based on the acoustic representation of sound for both types of stimuli using the Pearson product-moment correlation (cf. Khaligh-Razavi & Kriegeskorte, 2014; Kriegeskorte et al., 2008), thus yielding an RDM for each word and each sound on one side, and within the neural space on the other (Cichy et al., 2014; Kietzmann, Gert, Tong, & König, 2017).

Moreover, the role of the affective sound of words in general language processing can be investigated by the use of pattern classification algorithms (e.g., Mitchel et al., 2008). Focusing on the amygdala, for instance, results of different classifiers with different numbers of predictors could be compared, and the additional insight gained by each predictor could identify qualitative differences in the underlying processing. As a starting point, a first model relying solely on the affective meaning of stimuli could be complemented with adding a next predictor of affective sound in a second model, followed by the addition of their interaction in the third model. Comparing the accuracy of models would provide straightforward results on the importance of each factor. I would hypothesize significantly higher accuracy in the third model than in the first and second models.

In the neuroimaging studies reported in this dissertation, words were presented aurally. Therefore, the results obtained are only applicable to the visual domain to a limited extent, which means some questions related to visual word processing have been left open. For instance, it is still unknown whether the affective impact of phonemes is rooted in their acoustic features, which can only be encoded in the auditory cortex, or whether their articulatory features also play a role, as suggested by the Motor Theory of Speech (D'Ausilio et al., 2009; Liberman & Mattingly, 1985). Thus, the use of visually presented words is recommended for future research.

12.4.4 What shapes the basic affective tone of a text?

When moving beyond the lexical level of language, the literary genre of poetry seems to be of particular interest for the investigation of affective iconicity. In this work, I attempted to extend the methods and results of previous studies (see Chapter 4) to provide a more comprehensive picture of the role of sublexical features in shaping the overall affective meaning of a text. However, these studies only constitute a first step in investigating the *basic affective tone* of literary texts and poems, leaving room for improvements to be considered in future research.

To be able to make statements about the actual impact of the *basic affective tone* on the reader, the large number of influencing factors in the text, i.e., textual features, should be reduced to an extent that statistical inference can be possible. For this, an experimental setting with texts that have been systematically manipulated at the sublexical level of language is recommended. This will represent a challenge for future work due to the interconnectedness of the features in text and the interaction between different processing layers (Schrott and Jacobs, 2011; Jacobs 2015b). A guiding example for such manipulation was created by Bohrn et al. (2012a), in which proverbs were manipulated in different variants and the affective impact of each on the reader was measured. A possible example of such manipulation is the alternation of rhetorical features, such as rhyme, which lead to phonological recurrences (cf. Menninghaus et al., 2014).

As the statistical operationalization of the basic affective sound strongly relies on the use of a high quality linguistic corpus, the selection of an appropriate corpus for further investigations (for instance for English) is very important. In general, linguistic corpora serve as proxies for the mental representations which indicate how language is processed by normative or 'average' readers (Jacobs, 2018b). At the time of developing this method, SUBTLEX-DE, with 25 million words, was one of the most appropriate corpora (Brysbaert et al., 2011.) freely available for research goals. More recently, a number of linguistic corpora have become available, covering different needs for different research questions (see Jacobs, 2018b for a review). These include huge corpora of 1.5 billion words, e.g., dewac (Baroni, Bernardini, Ferraresi, & Zanchetta, 2009), to those specialized for empirical studies of poetry, i.e., GEPC (Gutenberg English Poetry Corpus; Jacobs, 2018), and GGPC (Gutenberg, German Poetry Corpus, Jacobs, in prep.).

As hypothesized by the NCPM and the proposed 4x4 matrix therein, different features (supralexical, interlexical, lexical and sublexical) at different relevant levels of textual analysis (metric, phonological, morpho-syntactic and semantic) can potentially influence aesthetic and affective processes of poetry reception. Due to this complexity and the extremely large number of relevant features at any of these levels, a combination of QNA (or Q2NA) tools (for feature extraction and machine learning methods) has been suggested (Jacobs, 2018b) in order to overcome the current issues. Although in an early stage, such a combination has been successfully implemented in recent investigations, e.g., in order to determine the relevant features influencing the ratings and the literariness of metaphors (Jacobs and Kinder, 2017, 2018), or to classify Shakespeare's sonnets (Jacobs et al., 2017; Abramo et al., 2018). At the sublexical level, this approach will provide a strong tool that can be used to extract more complex phonological features than those used in the present work (e.g., resonance, internal rhyme, or sonority-based euphony) and to examine their role in the aesthetic and affective responses of the reader at all possible levels of measurements, from ratings and peripheral-physiological responses all the way to activation patterns in the brain.

12.5 Conclusions

This dissertation, which was motivated by the long-lasting linguistic and philosophical debate on the standalone role of sublexical units (e.g., phonemes) in meaning making, investigated psychological mechanisms and neural networks underlying sound-to-meaning correspondence in the affective domain, termed affective iconicity. The empirical findings unveil a wide range of functions of the sound in words, from evoking affective responses as evident at both the psychological and neural level, influencing the processes of affective meaning making and semantic decisions, up to shaping the general affective meaning of poems as perceived by the readers.

The findings further illuminate which specific phonetic features of printed words and acoustic variables of spoken words contribute to the affective potential of the sound of words. This can effectively be used for constructing words and pseudowords associated with specific effects to be used in various contexts such as marketing, advertising, art, and literature. Thus, the present work suggests that single phonemes or phoneme clusters can be inherently meaningful, in the sense that they possess phonetic and acoustic features which can be related to affective experiences, thereby conveying meaning. This idea does not necessarily imply that pairings between form and meaning in language are non-arbitrary. Rather, the inherent meaningfulness of phonemes can potentially interact with the conventional meaning of words, and in the case of meaningless words (e.g., pseudowords) this property of phonemes can link the sound to semantic concepts.

A number of suggestions and extensions for the existing models of lexical word processing and supralexical literary reading were drawn from the results of the present work, which hopefully provide a theoretical framework for further investigation into the topic.

References

Abramo, F., Gambino, R., Pulvirenti, G., Shuwei, S., Sylvester, T., Mangen, A., Papp-Zipernovszky, O., Lüdtke, J. & Jacobs, A.M. (2018). A qualitative-quantitative analysis of Shakespeare's sonnet 60. Style, in revision.

Aggleton, J. P. (1993). The contribution of the amygdala to normal and abnormal emotional states. *Trends in Neurosciences*, *16*(8), 328–333.

Albers, S. (2008). *Lautsymbolik in ägyptischen Texten.* Mainz: Philipp Von Zabern Verlag Gmbh.

Altmann, U., Bohrn, I.C., Lubrich, O., Menninghaus, W., Jacobs, A.M. (2012b). The power of emotional valence—from cognitive to affective processes in reading. *Frontiers in human neuroscience* 01/2012; 6:192. doi: 10.3389/fnhum.2012.00192

Altmann, U., Bohrn, I.C., Lubrich, O., Menninghaus, W., Jacobs, A. M. (2012a). Fact versus Fiction - How paratextual information shapes our reading processes. *Social cognitive and affective neuroscience*. 2012; doi: 10.1093/scan/nss098

Anderson, A. K. (2005). Affective influences on the attentional dynamics supporting awareness. *Journal of Experimental Psychology: General*, *134*(2), 258–281. https://doi.org/10.1037/0096-3445.134.2.258

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, *33*(4), 245–266. https://doi.org/10.1017/S0140525X10000853

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating Experiential and Distributional Data to Learn Semantic Representations. *Psychological Review*, *116*(3), 463–498.

Aryani, A., & Jacobs, A. M. (2018b). Affective Congruence between Sound and Meaning of Words Facilitates Semantic Decision. *Behavioral Sciences (Basel, Switzerland)*, *8*(6).

Aryani, A., Conrad, M., Schmidtke, D., & Jacobs, A. (2018a). Why'piss' is ruder than'pee'? The role of sound in affective meaning making. *PloS one*, *13*(6), e0198430.

Aryani, A., Hsu, C.-T., & Jacobs, A. M. (2018c). The Sound of Words Evokes Affective Brain Responses. *Brain Sciences*, *8*(6). https://doi.org/10.3390/brainsci8060094

Aryani, A., Jacobs, A. M., & Conrad, M. (2013). Extracting salient sublexical units from written texts: "Emophon," a corpus-based approach to phonological iconicity. *Frontiers in Psychology*, *4*(OCT), 1–15. https://doi.org/10.3389/fpsyg.2013.00654

Aryani, A., Kraxenberger, M., Ullrich, S., Jacobs, A. M., & Conrad, M. (2016). Measuring the Basic Affective Tone of Poems via Phonological Saliency and Iconicity. *Psychology of Aesthetics , Creativity , and the Arts*, *Vol 10(2),* 191–204. https://doi.org/http://dx.doi.org/10.1037/ aca0000033

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage, 38*(1), 95–113. https://doi.org/10.1016/j.neuroimage.2007.07.007

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage, 26*(3), 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018

Astley, N. (Ed.). (2006). *Bloodaxe Poetry Introductions: Enzensberger, Holub, Sorescu, Tranströmer.* Highgreen: Bloodaxe Books.

Auracher, J., Albers, S., Zhai, Y., Gareeva, G., and Stavniychuk, T. (2011). P is for Happiness, N is for Sadness: universals in sound iconicity to detect emotions in poetry. *Discourse Processes*, *48*, 1–25. doi: 10.1080/01638531003674894

Aziz-Zadeh, L., & Damasio, A. (2008). Embodied semantics for actions: Findings from functional brain imaging. *Journal of Physiology-Paris*, *102*(1–3), 35–39.

Baayen, R. H., Piepenbrock, R., & van H, R. (1993). The ${$CELEX$}$ lexical data base on ${$CD-ROM$}$.

Bachorowski, J. A. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science*, *8*(2), 53-57.

Bachorowski, J. A., & Owren, M. J. (1995). Vocal expression of emotion: Acoustic Properties of Speech Are Associated With Emotional Intensity and Context. *Psychological Science*, *6*(4), 219–224.

Bailey, R. (1971). Statistics and the sounds of poetry. *Poetics*, 1, 16-37.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614.

Bänziger, T., & Scherer, K. R. (2005). The role of intonation in emotional expressions. *Speech Communication*, *46*(3), 252-267.

Bänziger, T., Hosoya, G., & Scherer, K. R. (2015). Path models of vocal emotion communication. *PLoS ONE*, *10*(9), 1–29. https://doi.org/10.1371/journal.pone.0136675

Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, *43*(3), 209–226.

Barrett, L. F. (2006). Are Emotions Natural Kinds? *Perspectives on Psychological Science*, *1*(1), 28–58.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, *59*, 617–645.

Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577-660.

Barthes, R. (1994). 11 The Death of the Author. *Media Texts, Authors and Readers: A Reader,* 166.

Baxter, S., & Lowrey, T. M. (2011). Phonetic symbolism and children's brand name preferences. *Journal of Consumer Marketing*, *28*(7), 516–523.

Belin, P., McAdams, S., Smith, B., Savel, S., Thivard, L., Samson, S., & Samson, Y. (1998). The functional anatomy of sound intensity discrimination. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *18*(16), 6388–6394. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9698330

Bergen, B. K. (2004). The psychological reality of phonaesthemes. *Language, 80*(2), 290–311.

Berlin, B. (1994). Evidence for pervasive synaesthetic sound symbolism in ethnozoological nomenclature. In L. Hinton, J. Nichols, & J. Ohala (Eds.), Cambridge University Press, *Sound symbolism*, pp. 77–93.

Berlin, B., Hinton, L., Nichols, J., & Ohala, J. J. (1994). Sound symbolism. *L. Hinton, J. Nichols, & J. Ohala (Eds.)*, 76-93.

Berthier, M. L., & Pulvermüller, F. (2011). Neuroscience insights improve neurorehabilitation of poststroke aphasia. *Nature Reviews Neurology*, *7*(2), 86.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhp055

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, *113*(39), 10818–10823.

Blomert, L. (2011). The neural signature of orthographic-phonological binding in successful and failing reading development. *NeuroImage*.

Boersma, P., & Weenik, D. (1996). PRAAT: a system for doing phonetics by computer. Report of the Institute of Phonetic Sciences of the University of Amsterdam. *Amsterdam: University of Amsterdam*.

Bohrn, I. C., Altmann, U., & Jacobs, A. M. (2012a). Looking at the brains behind figurative language-A quantitative meta-analysis of neuroimaging studies on metaphor, idiom, and irony processing. *Neuropsychologia*, *50*(11), 2669–2683.

Bohrn, I. C., Altmann, U., & Jacobs, A. M. (2012b). Looking at the brains behind figurative language-A quantitative meta-analysis of neuroimaging studies on metaphor, idiom, and irony processing. *Neuropsychologia*.

Bohrn, I. C., Altmann, U., Lubrich, O., Menninghaus, W., & Jacobs, A. M. (2013). When we like what we know - A parametric fMRI analysis of beauty and familiarity. *Brain and Language*, *124*(1), 1–8.

Bohrn, I. C., Altmann, U., Lubrich, O., Menninghaus, W., and Jacobs, A. M. (2012a). Old proverbs in new skins – an fMRI study on defamiliarization. *Frontiers in Psychology, 3*:204.

Boiten, F. A. (1998). The effects of emotional behaviour on components of the respiratory cycle. *Biological Psychology*, *49*(1–2), 29–51.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences*, *8*(12), 539–546.

Bowers, J. S., & Pleydell-Pearce, C. W. (2011). Swearing, euphemisms, and linguistic relativity. *PLoS ONE*, *6*(7).

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49-59.

Bradley, M.M., & Lang, P.J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (pp. 1-45). Technical Report C-1, The Center for Research in Psychophysiology, University of Florida

Brandt, L. (2010). Language and Enunciation – A Cognitive Inquiry with Special Focus on Conceptual Integration in Semiotic Meaning Construction. *Available at SSRN 1674588*

Braun, M., Hutzler, F., Ziegler, J. C., Dambacher, M., & Jacobs, A. M. (2009). Pseudohomophone effects provide evidence of early lexico-phonological processing in visual word recognition. *Human Brain Mapping*, *30*(7), 1977–1989.

Bredin, H. (1996). Onomatopoeia as a figure and a linguistic principle. *New Literary History*, 27(3), 555-569.

Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Linguistics and Language Compass*, *8*(2), 37–50.

Briesemeister, B. B., Kuchinke, L., & Jacobs, A. M. (2012). Emotional valence – A bipolar continuum or two independent dimensions? *SAGE Open*, 2(4).

Briesemeister, B. B., Kuchinke, L., & Jacobs, A. M. (2014). Emotion word recognition: Discrete information effects first, continuous later? *Brain Research*, *1564*, 62–71.

Briesemeister, B. B., Kuchinke, L., Jacobs, A. M., & Braun, M. (2015). Emotions in reading: Dissociation of happiness and positivity. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(2), 287–298.

Briesemeister, B. B., Kuchinke, L., Jacobs, A.M. & Braun, M. (2015). Emotions in reading: Dissociation of happiness and positivity. *Cognitive Affective & Behavioral Neuroscience*. In press.

Briesemeister, B.B., Kuchinke, L., & Jacobs, A.M. (2011a). Discrete Emotion Norms for Nouns - Berlin Affective Word List (DENN-BAWL). *Behavior Research Methods*, 43(2), 441-448.

Briesemeister, B.B., Kuchinke, L., & Jacobs, A.M. (2011b). Discrete Emotion Effects on Lexical Decision Response Times. *PLoS ONE*, 6(8), e23743.

Brück, C., Kreifelts, B., & Wildgruber, D. (2011). Emotional voices in context: A neurobiological model of multimodal affective information processing. *Physics of Life Reviews*, *8*(4), 383–403. https://doi.org/10.1016/j.plrev.2011.10.002

Brück, C., Kreifelts, B., Gößling-Arnold, C., Wertheimer, J., & Wildgruber, D. (2014). “Inner voices”: The cerebral representation of emotional voice cues described in literary texts. *Social Cognitive and Affective Neuroscience*, *9*(11), 1819–1827.

Brysbaert, M.; Buchmeier, M.; Conrad, M.; Jacobs, A. M.; Bölte, J.; Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, Vol 58(5), 2011, 412-424.

Bühler, K. (1934). *Sprachtheorie*. Oxford, England: Fischer. (1934).

Cai, Q., Brysbaert, M. (2010). SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles. *PLoS ONE* 5(6)

Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral Cortex (New York, N.Y. : 1991)*, *11*(12), 1110–1123.

Cannon, P. R., Hayes, A. E., & Tipper, S. P. (2010). Sensorimotor fluency influences affect: Evidence from electromyography. *Cognition & Emotion*, *24*(4), 681–691.

Carlin, G. (1990). *Doin’it Again*. MPI Media Group.

Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2013.11.005

Cassidy, K.W., Kelly, M.H., Sharoni, L.J. (1999). Inferring gender from name phonology. *Journal of Experimental Psychology: General*, 128, 362-381.

Chen, Y.-C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition*, *114*(3), 389–404.

Childs, G. T. (1994). African ideophones. *Sound Symbolism*. https://doi.org/10.1017/CBO9780511751806.013

Childs, G. T. (2015). Sound Symbolism, (Ching 1982), 1–6.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, 31(05), 489-509.

Christiansen, M. H., & Chater, N. (2015). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*.

https://doi.org/10.1017/S0140525X1500031X

Christiansen, M. H., & Kirby, S. (2003). Language evolution: Consensus and controversies. *Trends in Cognitive Sciences*, *7*(7), 300–307.

Christiansen, M. H., & Monaghan, P. (2016). Division of Labor in Vocabulary Structure: Insights From Corpus Analyses. *Topics in Cognitive Science*, *8*(3), 610–624.

Christiansen, M.H. & Monaghan, P. (in press). Division of labor in vocabulary structure: Insights from corpus analyses. *Topics in Cognitive Science*, in press

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience, 17*(3), 455.

Citron, F. M. M. (2012). Neural correlates of written emotion word processing: a review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and Language*, *122*(3), 211–226.

Citron, F. M. M., & Goldberg, A. E. (2014). Metaphorical sentences are more emotionally engaging than their literal counterparts. *Journal of Cognitive Neuroscience, 26*(11), 2585–2595.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. *Papers in Laboratory Phonology*, *1*, 283–333.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of Reading Aloud: Dual-Route and Parallel-Distributed-Processing Approaches. *Psychological Review*.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*.

Conrad, M., & Jacobs, A. M. (2004). Replicating syllable-frequency effects in Spanish in German: One more challenge to computational models of visual word recognition. *Language and Cognitive Processes, 19*(3), 369-390.

Conrad, M., Carreiras, M., Tamm, S., & Jacobs, A.M. (2009). Syllables and bigrams: Orthographic redundancy and syllabic units affect visual word recognition at different processing levels. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 461-479.

Conrad, M., Grainger, J., & Jacobs, A. M. (2007). Phonology as the source of syllable frequency effects in visual word recognition: Evidence from French. *Memory & Cognition, 35* (5), 974-983.

Conrad, M., Recio, G., & Jacobs, A. M. (2011). The Time Course of Emotion Effects in First and Second Language Processing: A Cross Cultural ERP Study with German–Spanish Bilinguals. *Frontiers in Psychology, 2,* 351.

Conrad, M., Tamm, S., Carreiras, M., & Jacobs, A. M. (2010). Simulating syllable frequency effects within an interactive activation framework. *European Journal of Cognitive Psychology*.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, *19*(5), 381–385.

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*.

Dambacher, M., Rolfs, M., Göllner, K., Kliegl, R., & Jacobs, A. M. (2009). Event-related potentials reveal rapid verification of predicted visual input. *PLoS ONE*.

Darwin, C. (1871). *The Descent of Man*. *Darwin*. https://doi.org/10.1016/S0140-6736(01)31142-X

Darwin, C. (1888). *The descent of man and selection in relation to sex* (Vol. 1). Murray.

De Carolis, L., Marsico, E., & Coupé, C. (2017). Evolutionary roots of sound symbolism. Association tasks of animal properties with phonetic features. *Language and Communication*, *54*, 21–35.

De Cuypere, L. (2008). *Limiting the iconic : from the metatheoretical foundations to the creative possibilities of iconicity in language*. John Benjamins Publ. Co.

de Ruiter, L. E., Theakston, A. L., Brandt, S., & Lieven, E. V. M. (2018). Iconicity affects children's comprehension of complex sentences: The role of semantics, clause order, input and individual differences. *Cognition*, *171*(March 2017), 202–224.

Deacon, T. (1997). *The Symbolic Species: The Co-evolution of Language and the Human Brain.* London: Allen Lane.

den Bergh, O., Vrana, S., & Eelen, P. (1990). Letters from the heart: Affective categorization of letter combinations in typists and nontypists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(6), 1153.

Derrida, J. (1976). Of grammatology, trans. *Gayatri Chakravorty Spivak (Baltimore, 1976)*, *141*.

Diffloth, G. (1972). Notes on expressive meaning. In *Chicago Linguistic Society* (Vol. 8, p. M4).

Dimitropoulou, M., Duñabeitia, J., Avilés, A., Corral, J. & Carreiras, M. (2010). Subtitle-based word frequencies as the best estimate of reading behaviour: the case of Greek. *Frontiers in Psychology*, 1:218, 1-12

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences*, *19*(10), 603–615.

Dingemanse, M., Schuerman, W., Reinisch, E., Tufvesson, S., & Mitterer, H. (2016). What sound symbolism can and cannot do: Testing the iconicity of ideophones from five languages. *Language*, *92*(2), e117--e133.

Dixon, P., Bortolussi, M. (2015). Measuring Literary Experience: Comment on Jacobs (2015). *Scientific Study of Literature*, 5(2), 178–182.

Doyle, J. R., & Bottomley, P. A. (2009). The massage in the medium: Transfer of connotative meaning from typeface to names and products. *Applied Cognitive Psychology*, 23, 396–409.

Doyle, J. R., & Bottomley, P. A. (2011). Mixed messages in brand names: Separating the impacts of letter shape from sound symbolism. *Psychology & Marketing*, 28(7), 749-762.

Dresler, T., Mériau, K., Heekeren, H. R., & Van Der Meer, E. (2009). Emotional Stroop task: Effect of word arousal and subject anxiety on emotional interference. *Psychological Research*, *73*(3), 364–371.

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior research methods*, 1-13.

Duddington, J. (2008). eSpeak 1.36. *See Http://Espeak. Sourceforge. Net*.

Durieux, G., & Gillis, S. (2001). Predicting grammatical classes from phonological cues: An empirical test. *LANGUAGE ACQUISITION AND LANGUAGE DISORDERS*, *23*, 189–230.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & der Vrecken, O. (1996). The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on* (Vol. 3, pp. 1393–1396).

Eco, U. (1976). *A theory of semiotics* (Vol. 217). Indiana University Press.

Elliott, E. A., & Jacobs, A. M. (2013). Facial expressions, emotions, and sign languages. *Frontiers in Psychology*, *4*, 115.

Elliott, E. A., & Jacobs, A. M. (2014). Phonological and morphological faces: Disgust signs in German Sign Language. *Sign Language & Linguistics*, *17*(2), 123–180. https://doi.org/10.1075/sll.17.2.01ell

Enfield, N. J. (2005). Areal linguistics and mainland Southeast Asia. *Annu. Rev. Anthropol.*, *34*, 181–206.

Enzensberger, H. M. (1981). *Verteidigung der Wölfe: Gedichte*. Suhrkamp.

Ethofer, T., Anders, S., Wiethoff, S., Erb, M., Herbert, C., Saur, R., … Wildgruber, D. (2006). Effects of prosodic emotional intensity on activation of associative auditory cortex. *Neuroreport*, *17*(3), 249–253.

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., & Vuilleumier, P. (2011). Emotional voice areas: anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. *Cerebral Cortex*, *22*(1), 191–200.

Ethofer, T., Van De Ville, D., Scherer, K., & Vuilleumier, P. (2009). Decoding of Emotional Information in Voice-Sensitive Cortices. *Current Biology*, *19*(12), 1028–1033.

Etkin, A., Egner, T., & Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in Cognitive Sciences*, *15*(2), 85–93.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences*, *103*(32), 12203–12208.

Fischer, O., & Nänny, M. (2001). *The motivated sign* (Vol. 2). John Benjamins Publishing.

Fitneva, S. A., Christiansen, M. H., & Monaghan, P. (2009). From sound to syntax: phonological constraints on children's lexical categorization of new words. *Journal of Child Language*, *36*(05), 967-997.

Flack, P. (2016). Roman Jakobson and the Transition of German Thought to the Structuralist Paradigm. *Acta Structuralica*, *1*, 1–15.

Fonagy, I. (1961). Communication in poetry. *Word, 17*, 194–218.

Fónagy, I. (2001). *Languages within language: An evolutive approach* (Vol. 13). John Benjamins Publishing.

Forgács, B., Bohrn, I., Baudewig, J., Hofmann, M. J., Pléh, C., & Jacobs, A. M. (2012). Neural correlates of combinatorial semantic processing of literal and figurative noun noun compound words. *NeuroImage*.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*. https://doi.org/10.1016/S0022-5371(73)80042-8

Foucault, M. (1982). The subject and power. *Critical Inquiry*, *8*(4), 777–795.

Freeman, M.H. (2009). Minding: Feeling, Form, and Meaning in the Creation of Poetic Iconicity (2009). *COGNITIVE POETICS: GOALS, GAINS, AND GAPS, Geert Brône and Jeroen Vandaele, (eds.),* pp. 169-196

Freeman, S., & Herron, J. C. (2007). *Evolutionary analysis*. Pearson Prentice Hall Upper Saddle River, New Jersey.

Frith, U. (1985). Beneath the surface of developmental dyslexia. *Surface Dyslexia*. https://doi.org/10.1016/j.jconhyd.2010.08.009

Frith, U. (1986). A developmental framework for developmental dyslexia. *Annals of Dyslexia*. https://doi.org/10.1007/BF02648022

Froehlich, E., Liebig, J., Morawetz, C., Ziegler, J. C., Braun, M., Heekeren, H. R., & Jacobs, A. M. (2018). Same Same But Different: Processing Words in the Aging Brain. *Neuroscience*. https://doi.org/10.1016/j.neuroscience.2017.11.042

Frühholz, S., & Grandjean, D. (2012). Towards a fronto-temporal neural network for the decoding of angry vocal expressions. *NeuroImage*, *62*(3), 1658–1666. https://doi.org/10.1016/j.neuroimage.2012.06.015

Frühholz, S., & Grandjean, D. (2013). Processing of emotional vocalizations in bilateral inferior frontal cortex. *Neuroscience and Biobehavioral Reviews*, *37*(10), 2847–2855.

Frühholz, S., Ceravolo, L., & Grandjean, D. (2011). Specific brain networks during explicit and implicit decoding of emotional prosody. *Cerebral Cortex*, *22*(5), 1107–1117.

Frühholz, S., Trost, W., & Grandjean, D. (2014). The role of the medial temporal limbic system in processing emotions in voice and music. *Progress in Neurobiology*, *123*, 1–17. https://doi.org/10.1016/j.pneurobio.2014.09.003

Frühholz, S., Trost, W., & Kotz, S. A. (2016). The sound of emotions-Towards a unifying neural network perspective of affective sound processing. *Neuroscience and Biobehavioral Reviews*, *68*(June), 1–15. https://doi.org/10.1016/j.neubiorev.2016.05.002

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology, 22*(3–4), 455–479.

Garvin, P. (1964). *A Prague School Reader on Esthetics, Literary Structures and Style*. Washington, DC: Georgetown University Press.

Gasser, M. (2004). The origins of arbitrariness in language. *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, 4–7. https://doi.org/DOI: 10.1098/rstb.2013.0299

Gavins, J., & Gerard S. (2003). *Cognitive Poetics in Practice*. London: Routledge.

Genette, G. (1976/1995). *Mimologics: Voyage en Cratylie*. Lincoln: University of Nebraska.

Glasser, M. F., & Rilling, J. K. (2008). DTI tractography of the human brain's language pathways. *Cerebral Cortex*, *18*(11), 2471–2482.

Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(4), 586–596.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565.

Glenberg, A. M., Webster, B. J., Mouilso, E., Havas, D., & Lindeman, L. M. (2009). Gender, emotion, and the embodiment of language comprehension. *Emotion Review, 1*(2), 151–161. https://doi.org/10.1177/1754073908100440

Glenberg, A.M. (1997). What memory is for. *Behavioral and Brain Sciences* (1997) 20, 1–55

Grainger, J., & Jacobs, a M. (1996). Orthographic processing in visual word recognition: a multiple read-out model. *Psychological Review*. https://doi.org/10.1037//0033-295X.103.3.518

Grainger, J., & Jacobs, A. M. (1999). Temporal integration of information in orthographic priming. *Visual Cognition*, *6*(3–4), 461–492.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, *8*(2), 145–146. https://doi.org/10.1038/nn1392

Greve, D. N. (2002). Optseq2. *Computer Software*.

Grimm, R. (1981); "Option für einen Klassiker" in: Enzensberger, H. M. : *die verteidigung der wölfe*. Frankfurt am Main: Suhrkamp Verlag, pp. 97-107.

Hakemulder, J. F. (2004). Foregrounding and its effect on readers' perception. *Discourse Processes*, *38*, 193–218. doi: 10.1207/s15326950dp3802_3

Hamano, S. (1998). *The Sound-Symbolic System of Japanese.* ERIC.

Hannagan, T., Amedi, A., Cohen, L., Dehaene-Lambertz, G., & Dehaene, S. (2015). Origins of the specialization for letters and numbers in ventral occipitotemporal cortex. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2015.05.006

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1–3), 335–346.

Hashimoto, T., Usui, N., Taira, M., Nose, I., Haji, T., & Kojima, S. (2006b). The neural mechanism associated with the processing of onomatopoeic sounds. *NeuroImage*, *31*(4), 1762–1770. https://doi.org/10.1016/j.neuroimage.2006.02.019

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic Representation of Action Words in Human Motor and Premotor Cortex. *Neuron*, *41*(2), 301–307. https://doi.org/10.1016/S0896-6273(03)00838-9

Heise, D. R. (1966). Sound-meaning correlations among 1,000 English words. *Language and Speech*, *9*(1), 14-27.

Hockett, C. F. (1958). *A course in modern linguistics*. Oxford And Ibh Publishing Co.; New Delhi.

Hockett, C. F. (1960). *A Course in Modern Linguistics*. London: Macmillan.

Hockett, C. F. (1977). *The view from language: Selected essays, 1948-1974*. University of Georgia Press.

Hofmann, M. J., & Jacobs, A. M. (2014). Interactive activation and competition models and semantic context: From behavioral to brain data. *Neuroscience and Biobehavioral Reviews*, *46*(P1), 85–104. https://doi.org/10.1016/j.neubiorev.2014.06.011

Hofmann, M. J., Herrmann, M. J., Dan, I., Obrig, H., Conrad, M., Kuchinke, L., … Fallgatter, A. J. (2008). Differential activation of frontal and parietal regions during visual word recognition: An optical topography study. *NeuroImage*, *40*(3), 1340–1349. https://doi.org/10.1016/j.neuroimage.2007.12.037

Hofmann, M. J., Kuchinke, L., Tamm, S., Võ, M. L. H., & Jacobs, A. M. (2009). Affective processing within 1/10th of a second: High arousal is necessary for early facilitative processing of negative but not positive words. *Cognitive, Affective and Behavioral Neuroscience*, *9*(4), 389–397. https://doi.org/10.3758/9.4.389

Hofmann, M. J., Stenneken, P., Conrad, M., & Jacobs, A. M. (2007). Sublexical frequency measures for orthographic and phonological units in German. *Behavior Research Methods*, *39*(3), 620–629. https://doi.org/10.3758/BF03193034

Hofmann, M. J., Tamm, S., Braun, M. M., Dambacher, M., Hahne, A., & Jacobs, A. M. (2008). Conflict monitoring engages the mediofrontal cortex during nonword processing. *NeuroReport*, *19*(1), 25–29. https://doi.org/10.1097/WNR.0b013e3282f3b134

Holmes, A. P., & Friston, K. J. (1998). Generalisability, random Effects & population inference. *Neuroimage*, *7*, S754.

Hsu, C. T., Conrad, M., & Jacobs, A. M. (2014). Fiction feelings in Harry Potter: Haemodynamic response in the mid-cingulate cortex correlates with immersive reading experience. *NeuroReport*. https://doi.org/10.1097/WNR.0000000000000272

Hsu, C. T., Jacobs, A. M., Citron, F. M. M., & Conrad, M. (2015a). The emotion potential of words and passages in reading Harry Potter - An fMRI study. *Brain and Language*. https://doi.org/10.1016/j.bandl.2015.01.011

Hsu, C.T., Conrad, M., & Jacobs, A.M. (2013). Why Should German Bilinguals Read Harry Potter in German? Paper given at the 55th Conference of Experimental Psychologists, Vienna, March 2013.

Hsu, C.T., Jacobs, A.M., Altmann, U., & Conrad, M. (2015b). The magical activation of left amygdala when reading Harry Potter: An fMRI study on how descriptions of supra-natural events entertain and enchant. *Plos One*, *10* (2), e0118179

Huang, Y., Pratoomraj, S., & Johnson, R. C. (1969). Universal magnitude symbolism. *Journal of Verbal Learning & Verbal Behavior*, *8*, 155–156. doi: 10.1016/S0022-5371(69)80028-9

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130298–20130298. https://doi.org/10.1098/rstb.2013.0298

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, *109*(1), 54–65. https://doi.org/10.1016/j.cognition.2008.07.015

Iser, W. (1994). *Der Akt des Lesens: Theorie ästhetischer Wirkung.* . Stuttgart: UTB.

Iwasaki, N., Vinson, D. P., & Vigliocco, G. (2007a). How does it hurt, kiri-kiri or siku-siku? Japanese mimetic words of pain perceived by Japanese speakers and English speakers.

Iwasaki, N., Vinson, D. P., & Vigliocco, G. (2007b). What do English Speakers Know about gera-gera and yota-yota？: A Cross-linguistic Investigation of Mimetic Words for Laughing and Walking. *Japanese-Language Education around the Globe*, *17*(6), 53–78.

Jacobs, A. M. (2011). Neurokognitive Poetik: Elemente eines Modells des literarischen Lesens (neurocognitive poetics: elements of a model of literary reading). I n Schrott & A. M. Jacobs (Eds.), *Gehirn und Gedicht: Wie wir Unsere Wirklichkeiten Konstruieren*, eds R. (pp. 492–520), München: Carl Hanser.

Jacobs, A. M. (2015a). Neurocognitive poetics: methods and models for investigating the neuronal and cognitive-affective bases of literature reception. *Frontiers in Human Neuroscience*, *9*(April), 1–22. https://doi.org/10.3389/fnhum.2015.00186

Jacobs, A. M. (2015b). "Towards a neurocognitive poetics model of literary reading," in *Towards a Cognitive Neuroscience of Natural Language Use*, ed R. Willems (Cambridge: Cambridge University Press), 135–159.

Jacobs, A. M. (2017). Quantifying the Beauty of Words: A Neurocognitive Poetics Perspective. *Frontiers in Human Neuroscience*, *11*(December), 1–7. https://doi.org/10.3389/fnhum.2017.00622

Jacobs, A. M. (2018a). The Gutenberg English Poetry Corpus: Exemplary Quantitative Narrative Analyses . *Frontiers in Digital Humanities* .

Jacobs, A. M. (2018b). Computational Stylistics and (Neuro-)Cognitive Poetics, *Scientific Study of Literature*

Jacobs, A. M., & Grainger, J. (1994). Models of visual word recognition: Sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037//0096-1523.20.6.1311

Jacobs, A. M., & Hofmann, M. J. (2011). Neurokognitive Modellierung Neurocognitive modeling Arthur M. Jacobs, Markus J. Hofmann, 1–18.

Jacobs, A. M., & Kinder, A. (2017). "The Brain Is the Prisoner of Thought": A Machine-Learning Assisted Quantitative Narrative Analysis of Literary Metaphors for Use in Neurocognitive Poetics. *Metaphor and Symbol*. https://doi.org/10.1080/10926488.2017.1338015

Jacobs, A. M., & Kinder, A. (2018). What makes a metaphor literary? Answers from two computational studies. *Metaphor and Symbol*.

Jacobs, A. M., & Lüdtke, J. (2017). Immersion into narrative and poetic worlds. *Narrative Absorption*, *27*, 69.

Jacobs, A. M., & Schrott, R. (2015). Captivated by the cinema of the mind. In *Concentration, Fiktion. CC* (pp. 118–149).

Jacobs, A. M., & Willems, R. M. (2017). The Fictive Brain: Neurocognitive Correlates of Engagement in Literature. *Review of General Psychology*. https://doi.org/10.1037/gpr0000106

Jacobs, A. M., & Ziegler, J. C. (2015). Visual Word Recognition, Neurocognitive Psychology of. In *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*. https://doi.org/10.1016/B978-0-08-097086-8.57021-4

Jacobs, A. M., and Kinder, A. (2015) 'Worte als Worte erfahren: wie erarbeitet das Gehirn Gedichte in *Kind und Gedicht,* ed A. Pompe (Berlin: Rombach).

Jacobs, A. M., Grainger, J., & Ferrand, L. (1995). The incremental priming technique: A method for determining within-condition priming effects. *Perception & Psychophysics*, *57*(8), 1101–1110.

Jacobs, A. M., Hofmann, M. J., & Kinder, A. (2016). On elementary affective decisions: To like or not to like, that is the question. *Frontiers in Psychology*, *7*(NOV), 1–17. https://doi.org/10.3389/fpsyg.2016.01836

Jacobs, A. M., Lüdtke, J., Aryani, A., Meyer-Sickendieck, B., & Conrad, M. (2016). Mood-empathic and aesthetic responses in poetry reception. *Scientific Study of Literature*, *6*(1), 87–130.

Jacobs, A. M., Rey, A., Ziegler, J. C., & Grainger, J. (1998). MROM-P: An interactive activation, multiple read-out model of orthographic and phonological processes in visual word recognition. In J. Grainger, J. & A.M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 147-187). Mahwah, NJ: Erlbaum

Jacobs, A. M., Schuster, S., Xue, S., & Lüdtke, J. (2017). What's in the brain that ink may character…. *Scientific Study of Literature*, *7*(1), 4–51.

Jacobs, A. M., Võ, M. L. H., Briesemeister, B. B., Conrad, M., Hofmann, M. J., Kuchinke, L., … Braun, M. (2015). 10 years of BAWLing into affective and aesthetic processes in reading: What are the echoes? *Frontiers in Psychology*, *6*(JUN), 1–15. https://doi.org/10.3389/fpsyg.2015.00714

Jacobs, A.M. (2013). Immersion and aesthetic trajectory in literary reading: what are the underlying neurocognitive processes? Talk given at the 'Lost in storyworlds symposium' (F. Hakemulder, E. Tan, M. Kuijpers), Utrecht, March 2013. (http://blog.hum.uu.nl/storycenter/)

Jacobs, A.M. (2014). Towards a Neurocognitive Poetics: development and tests of a dual-route model of literary reading. In R. Willems (ed.), *Towards a cognitive neuroscience of natural language use*. Cambridge Univ. Press.

Jakobson, R. (1937). *Lectures on Sound and Meaning*. MIT Press, Cambridge, MA.

Jakobson, R. (1960). Closing statement: Linguistics and poetics. In T. A. Sebeok (Ed.), *Style in Language,* Cambridge, MA: MIT Press. pp. 350–377.

Jakobson, R. (1965). Quest for the Essence of Language. *Diogenes*, *13*(51), 21–37. https://doi.org/10.1177/039219216501305103

Jakobson, R. (1966). Quest for the Essence of Language, *Selected Writings II*, 51: P.36.

Jakobson, R. (1970). On the verbal art of William Blake and other poet-painters. *Linguistic Inquiry*, *1*(1), 3-23.

Jakobson, R., & Waugh, L. R. (1979/ 2002). *The Sound Shape of Language*. Berlin: Walter de Gruyter.

Janata, P., Tomic, S. T., & Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. *Journal of Experimental Psychology: General*, *141*(1), 54–75. https://doi.org/10.1037/a0024208

Janssen, D. P. (2012). Twice random, once mixed: Applying mixed models to simultaneously analyze random effects of language and participants. *Behavior Research Methods*, *44*(1), 232–247. https://doi.org/10.3758/s13428-011-0145-1

Jespersen, O. (1922). *Language - its Nature, Development and Origin*. George Allen & Unwim Ltd., London.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260.

Jürgens, U. (2003). From mouth to mouth and hand to hand: On language evolution. *Behavioral and Brain Sciences*, *26*(2), 229–230.

Juslin, P. N., & Laukka, P. (2003). Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code? *Psychological Bulletin*, *129*(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Kanero, J., Imai, M., Okuda, J., Okada, H., & Matsuda, T. (2014). How sound symbolism is processed in the brain: a study on Japanese mimetic words. *PLoS One*, *9*(5), e97905.

Kanske, P., & Kotz, S. A. (2011). Emotion triggers executive attention: Anterior cingulate cortex and amygdala responses to emotional words in a conflict task. *Human Brain Mapping*, *32*(2), 198–208. https://doi.org/10.1002/hbm.21012

Kantartzis, K., Kita, S., & Imai, M. (2011). Japanese sound symbolism facilitates word learning in English speaking children. *Cognitive Science*, *35*(3), 575–586.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, *99*(2), 349.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control.

Science, *303*(5660), 1023–1026.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633. https://doi.org/10.3758/BRM.42.3.627

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, 42(3), 643-650.

Khaligh-Razavi, S.-M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology*, *10*(11), e1003915.

Kietzmann, T. C., Gert, A. L., Tong, F., & König, P. (2017). Representational dynamics of facial viewpoint encoding. *Journal of Cognitive Neuroscience*, *29*(4), 637–651.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*. https://doi.org/10.1037/0033-295X.85.5.363

Kirsner, R. S. (1985). Iconicity and grammatical meaning. *Iconicity in Syntax*, 249–270.

Kita, S. (1997). Two-dimensional semantic analysis of Japanese mimetics. Walter de Gruyter, Berlin/New York.

Klasen, M., Chen, Y. H., & Mathiak, K. (2012). Multisensory emotions: Perception, combination and underlying neural processes. *Reviews in the Neurosciences*, *23*(4), 381–392. https://doi.org/10.1515/revneuro-2012-0040

Klasen, M., Kenworthy, C. A., Mathiak, K. A., Kircher, T. T. J., & Mathiak, K. (2011). Supramodal Representation of Emotions. *Journal of Neuroscience*, *31*(38), 13635–13643. https://doi.org/10.1523/JNEUROSCI.2833-11.2011

Koelsch, S., Jacobs, A. M., Menninghaus, W., Liebal, K., Klann-Delius, G., von Scheve, C., & Gebauer, G. (2015). The quartet theory of human emotions: An integrative and neurofunctional model. *Physics of Life Reviews*. https://doi.org/10.1016/j.plrev.2015.03.001

Koelsch, S., Skouras, S., Fritz, T., Herrera, P., Bonhage, C., Küssner, M. B., & Jacobs, A. M. (2013). The roles of superficial amygdala and auditory cortex in music-evoked fear and joy. *NeuroImage*, *81*, 49–60. https://doi.org/10.1016/j.neuroimage.2013.05.008

Koffka, K. (1935). *Principles of Gestalt psychology*. London: Routledge and Kegan Paul.

Köhler, W. (1929). *Gestalt psychology*. New York: Liveright.

Köhler, W. (1947). *Gestalt psychology* (2nd edn.). New York: Liveright Publishin

Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends in Cognitive Sciences*, *14*(9), 392–399.

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The Representation of Abstract Words: Why Emotion Matters. *Journal of Experimental Psychology: General*, *140*(1), 14–34. https://doi.org/10.1037/a0021446

Kousta, S. T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, *112*(3), 473–481. https://doi.org/10.1016/j.cognition.2009.06.007

Kraxenberger, M. (2014). Jakobson Revisited: Poetic Distinctiveness, Modes of Operation, and Perception. *Rivista Italiana di Filosofia del Linguaggio*, *8*(1).

Kraxenberger, M., & Menninghaus, W. (2016). Mimological reveries? Disconfirming the hypothesis of phono-emotional iconicity in poetry. *Frontiers in Psychology*, *7*(NOV), 1–9. https://doi.org/10.3389/fpsyg.2016.01779

Kreifelts, B., Ethofer, T., Huberle, E., Grodd, W., & Wildgruber, D. (2010). Association of trait emotional intelligence and individual fMRI-activation patterns during the perception of social signals from voice and face. *Human Brain Mapping*, *31*(7), 979–991. https://doi.org/10.1002/hbm.20913

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4.

Kuchinke, L., Jacobs, A. M., Grubich, C., Võ, M. L. H., Conrad, M., & Herrmann, M. (2005). Incidental effects of emotional valence in single word processing: an fMRI study. *Neuroimage*, *28*(4), 1022-1032.

Kuchinke, L., Jacobs, A. M., Grubich, C., Võ, M. L.-H., Conrad, M., & Herrmann, M. (2005). Incidental effects of emotional valence in single word processing: An fMRI study. *NeuroImage*, 28(4), 1022-1032.

Kuchinke, L., Jacobs, A. M., Võ, M. L.-H., Conrad, M., Grubich, C., & Herrmann, M. (2006). Modulation of prefrontal cortex activation by emotional words in recognition memory. *NeuroReport*, 17(10), 1037-1041.

Kuhlmann, M., Hofmann, M. J., Briesemeister, B. B., & Jacobs, A. M. (2016). Mixing positive and negative valence: Affective-semantic integration of bivalent words. *Scientific Reports*, *6*, 1–7. https://doi.org/10.1038/srep30718

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*. https://doi.org/10.1038/307161a0

Kwon, N., & Round, E. R. (2014). Phonaesthemes in morphological theory. *Morphology*, *25*(1), 1–27. https://doi.org/10.1007/s11525-014-9250-z

Laing, C. E. (2014). A phonological analysis of onomatopoeia in early word production. *First Language*, *34*(5), 387–405.

Laing, C. E. (2015). What does the cow say ?' An exploratory analysis of development, (October), 1–271.

Lakoff, G., & Turner, M. (1989). More than cool reason: The power of poetic metaphor. *Berkeley: University of California at Berkeley*.

Lamb, S. M. (1964). The Sememic Approach to Structural Semantics1. *American Anthropologist*, *66*(3), 57-78.

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., … Fox, P. T. (2000). Automated Talairach Atlas labels for functional brain mapping. *Human Brain Mapping*, *10*(3), 120–131. https://doi.org/10.1002/1097-0193(200007)10:3<120::AID-HBM30>3.0.CO;2-8

Lang, P. J. (1995). The emotion probe: studies of motivation and attention. *American Psychologist*, *50*(5), 372.

Langer, S.K. (1953). *Feeling and Form: A Theory of Art 'Developed from Philosophy in a New Key.'* New York: Charles Scribner's.

Leech, G. (2008). *Language in Literature: Style and Foregrounding*, Harlow, Pearson/Longman Education.

Lehmann, C. (2007). Motivation in language. Attempt at a systematization. *Sprachliche Motivation. Zur Interdependenz von Inhalt Und Ausdruck*, 100–135.

Lehne, M., Engel, P., Rohrmeier, M., Menninghaus, W., Jacobs, A. M., & Koelsch, S.

(2015). Reading a suspenseful literary text activates brain areas related to social cognition and predictive inference. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0124550

Leiner, D. J. (2014). SoSci survey (version 2.4. 00-i)[computer software]. *Accessed February, 20,* 2015.

Leitman, D. I., Edgar, C., Berman, J., Gamez, K., Fruhholz, S., & Roberts, T. P. L. (2016). Amygdala and insula contributions to dorsal-ventral pathway integration in the prosodic neural network, (215), 1–12. Retrieved from http://arxiv.org/abs/1611.01643

Leitman, D. I., Wolf, D. H., Ragland, J. D., Laukka, P., Loughead, J., Valdez, J. N., … Gur, R. (2010). " It's not what you say, but how you say it": a reciprocal temporo-frontal network for affective prosody. *Frontiers in Human Neuroscience*, *4,* 19.

Lévi-Strauss, C. (1960). On manipulated sociological models. *Bijdragen Tot de Taal-, Land-En Volkenkunde*, (1ste Afl), 45–54.

Lewis, P. A., Critchley, H. D., Rotshtein, P., & Dolan, R. J. (2007). Neural correlates of processing valence and arousal in affective words. *Cerebral Cortex*, *17*(3), 742–748. https://doi.org/10.1093/cercor/bhk024

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1-36.

Lieberman, M. D., & Cunningham, W. A. (2009). Type I and Type II error concerns in fMRI research: Re-balancing the scale. *Social Cognitive and Affective Neuroscience*, *4*(4), 423–428. https://doi.org/10.1093/scan/nsp052

Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology*, *6*(OCT), 1–14. https://doi.org/10.3389/fpsyg.2015.01624

Lockwood, G., Hagoort, P., & Dingemanse, M. (2016). How iconicity helps people learn new words: Neural correlates and individual differences in sound-symbolic bootstrapping. *Collabra: Psychology*, *2*(1).

Löfberg, O., Julkunen, P., Pääkkönen, A., & Karhu, J. (2014). The auditory-evoked arousal modulates motor cortex excitability. *Neuroscience*, *274*, 403–408.

Louwerse, M.M. (2007). Symbolic or embodied representations: A case for symbol interdependency. In T. Landauer, D. McNamara, S. Dennis, & W.

Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 107–120). Mahwah, NJ: Erlbaum.

Louwerse, M.M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, 3(2), 273-302.

Louwerse, M.M., & Jeuniaux, P. (2008). Language comprehension is both embodied and symbolic. In M. de Vega, A. Glenberg, & A. C. Graesser (Eds.), *Embodiment and meaning: A debate* (pp. 309–326). Oxford, England: Oxford University Press.

Lüdtke, J., Meyer-Sickendieck, B., & Jacobs, A. M. (2014). Immersing in the stillness of an early morning: Testing the mood empathy hypothesis of poetry reception. *Psychology of Aesthetics, Creativity, and the Arts, 8*(3), 363.

Ma, W., & Thompson, W. F. (2015). Human emotions track changes in the acoustic environment. *Proceedings of the National Academy of Sciences*, *112*(47), 14563–14568. https://doi.org/10.1073/pnas.1515087112

MacLean, P. D. (1990). *The triune brain in evolution: Role in paleocerebral functions*. Springer Science & Business Media.

Magnus, M. (2001). *What's in a word? Studies in phonosemantics*. PhD thesis. Department of Linguistics, Norwegian University.

Majid, A. (2012). Current emotion research in the language sciences. *Emotion Review*, *4*(4), 432-443.

Mandelker, A. (1983). Russian Formalism and the Objective Analysis of Sound in Poetry, *The Slavic and East European Journal*, Vol. 27, No. 3, pp. 327-338

Marlin, D. (2008). Saussure and the elusive question of the origin. *Semiotica, 2008*(172), 185–200.

Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research*. https://doi.org/10.1007/BF01067101

Masson, D. I. (1954). Free phonetic patterns in Shakespeare's sonnets. *Neophilologus, 38*(1), 277–289.

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science*, *9*(3), 316–322. https://doi.org/10.1111/j.1467-7687.2006.00495.x

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*. https://doi.org/10.1037/0033-295X.88.5.375

McLaren, D. G., Ries, M. L., Xu, G., & Johnson, S. C. (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. *Neuroimage*, *61*(4), 1277–1286.

Mehrabian, A. (1980). Basic dimensions for a general psychological theory implications for personality, social, environmental, and developmental studies.

Menninghaus, W., Bohrn, I. C., Altmann, U., Lubrich, O., & Jacobs, A. M. (2014). Sounds funny? Humor effects of phonological and prosodic figures of speech. *Psychology of Aesthetics, Creativity, and the Arts*, *8*(1), 71–76. https://doi.org/10.1037/a0035309

Menninghaus, W., Bohrn, I. C., Altmann, U., Lubrich, O., & Jacobs, A. M. (2014). Sounds funny? Humor effects of phonological and prosodic figures of speech. *Psychology of Aesthetics, Creativity, and the Arts*, *8*, 71–76.

Menninghaus, W., Bohrn, I. C., Knoop, C. A., Kotz, S. A., Schlotz, W., & Jacobs, A. M. (2015). Rhetorical features facilitate prosodic processing while handicapping ease of semantic comprehension. *Cognition, 143,* 48-60. doi:10.1016/j.cognition.2015.05.026

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*(6174), 1006–1010. https://doi.org/10.1126/science.1245994

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788–804. https://doi.org/10.1016/j.cortex.2010.11.002

Meteyard, L., Stoppard, E., Snudden, D., Cappa, S. F., & Vigliocco, G. (2015). When semantics aids phonology: A processing advantage for iconic word forms in aphasia. *Neuropsychologia*, *76*, 264–275. https://doi.org/10.1016/j.neuropsychologia.2015.01.042

Meyer-Sickendiek, B. (2011). *Lyrisches Gespür: Vom geheimen Sensorium moderner Poesie.*. Paderborn, Germany: Fink.

Miall, D. S. (1995). Anticipation and feeling in literary response: A neuropsychological perspective. *Poetics*, *23*(4), 275-298.

Miall, D. S. (2001). Sounds of contrast: An empirical approach to phonemic iconicity. *Poetics*, *29*(1), 55-70.

Miall, D.S., & Kuiken, D. (1994). Foregrounding, defamiliarization, and affect response to literary stories. *Poetics* 22, 389–407.

Mirz, F., Gjedde, A., Sødkilde-Jrgensen, H., & Pedersen, C. B. (2000). Functional brain imaging of tinnitus-like perception induced by aversive auditory stimuli. *NeuroReport*, *11*(3), 633–637. https://doi.org/10.1097/00001756-200002280-00039

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, *55*(4), 259–305. https://doi.org/10.1016/j.cogpsych.2006.12.001

Monaghan, P., Christiansen, M.H., Fitneva, S.A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General*, Vol 140(3), Aug 2011, 325-347.

Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130299–20130299. https://doi.org/10.1098/rstb.2013.0299

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist*, *111*(981), 855–869.

Mukarovsky, J. (1964). “Standard language and poetic language” , in A *Prague School Reader on Esthetics, Literary Structures and Style*, ed P. Garvin (Washington, DC: Georgetown University Press), 17-30.

Muller, F. M. (1879). *Upanishads Part-1*. Motilal Banarsidass; Delhi.

Müller, V. I., Cieslik, E. C., Turetsky, B. I., & Eickhoff, S. B. (2012). Crossmodal interactions in audiovisual emotion processing. *NeuroImage*, *60*(1), 553–561. https://doi.org/10.1016/j.neuroimage.2011.12.007

Myers-Schulz, B., Pujara, M., Wolf, R.C. & Koenigs, M. (2013). Inherent emotional quality of human speech sounds, *Cognition & Emotion*, DOI:10.1080/02699931.2012.754739

Nastase, V., Sokolova, M. and Sayyad Shirabad, J. (2007). Do happy words sound happy? A study of the relation between form and meaning for English words expressing emotions. *In Proceedings of the Recent Advances in Natural Language Processing*, 406-410.

Newman, S. (1933). Further Experiments in Phonetic Symbolism. *American Journal of Psychology*. 45: 53.

Nicklas, P., & Jacobs, A. M. (2017). Rhetoric, Neurocognitive Poetics, and the Aesthetics of Adaptation. *Poetics Today*. https://doi.org/10.1215/03335372-3869311

Nöth, W. (2001). "Semiotic Foundations of Iconicity in Language and Literature", in *Iconicity in Language and Literature*, edited by Max Nänny and Olga Fischer, Amsterdam/Philadelphia: Benjamins

Nuerk, H. C., Rey, A., Graf, R., & Jacobs, A. M. (2000). Phonographic sublexical units in visual word recognition. *Current Psychology Letters*, (2), 25.

Nygaard, L.C., Cook, A.E., and Namy, L.L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition* 112(1), 181-186.

Nygaard, L.C., Herold, D.S., Namy, L.L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science*, 33, 127-146.

Nyklíček, I., Thayer, J. F., & Van Doornen, L. J. P. (1997). Cardiorespiratory differentiation of musically-induced emotions. *Journal of Psychophysiology*.

Obermeier, C., Menninghaus, W., Koppenfels, M. von, Raettig, T., Schmidt-Kassow, M., Otterbein, S., & Kotz, S. A. (2013). Aesthetic and Emotional Effects of Meter and Rhyme in Poetry. *Frontiers in Psychology*, *4*.

Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In Hinton, L., Nichols, J. & Ohala, J.J. (Eds.), *Sound Symbolism* (pp. 325–347) Cambridge: Cambridge University Press.

Ohala, J. J. (1996). Ethological theory and the expression of emotion in the voice. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*, *3*(November 1996), 1812–1815. https://doi.org/10.1109/ICSLP.1996.607982

Ohala, J. J. (1997). Sound Symbolism. *4th Seoul International Conference on Linguistics [SICOL]*, (Ching 1982), 98–103. https://doi.org/10.1525/jlin.1996.6.1.109

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.

Osaka, M., Osaka, N., Kondo, H., Morishita, M., Fukuyama, H., Aso, T., & Shibasaki, H. (2003). The neural basis of individual differences in working memory capacity: an fMRI study. *NeuroImage*, *18*(3), 789–797.

Osaka, N., Osaka, M., Morishita, M., Kondo, H., & Fukuyama, H. (2004). A word expressing affective pain activates the anterior cingulate cortex in the human brain: an fMRI study. *Behavioural Brain Research*, *153*(1), 123–127.

Osgood, C. E. (1952). The nature and measurement of meaning. *Psychological Bulletin*, *49*(3), 197.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). The measurement of meaning. Urbana: Univer. of Illinois Press, 195, 36.

Otis, K., and Sagi, E. (2008). Phonaesthemes: A corpora-base analysis. In B. C. Love, K. McRae, & V. M. Sloutsk (Eds.), *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*.

Ozturk, O., Krehm, M. , Vouloumanos, A. (2013) Sound symbolism in infancy: Evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*, Volume 114, issue 2

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. *Affective Science*. https://doi.org/10.1017/CBO9781107415324.004

Panksepp, J. (2000). The riddle of laughter: Neural and psychoevolutionary underpinnings of joy. *Current Directions in Psychological Science*, *9*(6), 183–186.

Panksepp, J. (2010). Emotional causes and consequences of social-affective vocalization. In *Handbook of Behavioral Neuroscience* (Vol. 19, pp. 201–208). Elsevier.

Parise, C. V, & Pavani, F. (2011). Evidence of sound symbolism in simple vocalizations. *Experimental Brain Research*, *214*(3), 373–380.

Peirce, C. S. (1931). *The Collected Papers*, eds C. Hartshorne, P. Weiss, and A. Burks, Cambridge: Harvard University Press.

Peña, M., Mehler, J., & Nespor, M. (2011). The role of audiovisual processing in early conceptual development. *Psychological Science*, *22*(11), 1419–1421. https://doi.org/10.1177/0956797611421791

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2013.0300

Perniss, P., Thompson, R.L. and Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Front. Psychology* 1:227. doi: 10.3389/fpsyg.2010.00227

Perrone-Bertolotti, M., Kujala, J., Vidal, J. R., Hamame, C. M., Ossandon, T., Bertrand, O., … Lachaux, J.-P. (2012). How Silent Is Silent Reading? Intracerebral Evidence for Top-Down Activation of Temporal Voice Areas during Reading. *Journal of Neuroscience*, *32*(49), 17554–17562. https://doi.org/10.1523/JNEUROSCI.2982-12.2012

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*. https://doi.org/10.1037/0033-295X.114.2.273

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: from a'low road'to'many roads' of evaluating biological significance. *Nature Reviews Neuroscience, 11*(11), 773.

Plato (1892). *"Cratylus" in the Dialogues of Plato*, Vol. 1, trans. Benjamin Jowett, Oxford: Clarendon, 253–89.

Ploog, D. (2002). Is the neural basis of vocalisation different in non-human primates and Homo sapiens? *The Speciation of Modern Homo Sapiens*, 121–135. https://doi.org/10.5871/bacad/9780197263112.003.0007

Ploog, D. W. (1992). The evolution of vocal communication. *Nonverbal Vocal Communication: Comparative and Developmental Approaches*, *6*.

Poe, E. A. (1881). *The bells*. Philadelphia: Porter & Coates.

Ponz, A., Montant, M., Liegeois-Chauvel, C., Silva, C., Braun, M., Jacobs, A. M., & Ziegler, J. C. (2014). Emotion processing in words: A test of the neural re-use hypothesis using surface and intracranial EEG. *Social Cognitive and Affective Neuroscience*. https://doi.org/10.1093/scan/nst034

Price, C. J. (2012). A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2012.04.062

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, 22, 253-336

Ramachandran, V. S., & Hubbard, E. H. (2001). Synaesthesia—A Window Into Perception, Thought and Language. *Journal of Consciousness Studies*, *8*(12), 3–34. https://doi.org/10.1111/1468-0068.00363

Reber, R., Schwarz, N., Winkielman, P., Schwarz We thank Marilyn Brewer, N., Fazendeiro, T., Feist, G., … Wittenbrink, B. (2004). Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience? *Personality and Social Psychology Review*, *8*(4), 364–382.

Recio, G., Conrad, M., Hansen, L. B., & Jacobs, A. M. (2014). On pleasure and thrill: The interplay between arousal and valence during visual word recognition. *Brain and Language*, *134*, 34–43. https://doi.org/10.1016/j.bandl.2014.03.009

Reilly, J., Westbury, C., Kean, J., & Peelle, J. E. (2012). Arbitrary symbolism in natural language revisited: When word forms carry meaning. *PLoS ONE*, *7*(8). https://doi.org/10.1371/journal.pone.0042286

Reuven, T. (1997). Sound Affects of Poetry—Critical Impressionism, Reductionism, and Cognitive Poetics. *Pragmatics and Cognition* 5: 283-304.

Revill, K. P., Namy, L. L., DeFife, L. C., & Nygaard, L. C. (2014). Cross-linguistic sound symbolism and crossmodal correspondence: Evidence from fMRI and DTI. *Brain and Language*, *128*(1), 18–24. https://doi.org/10.1016/j.bandl.2013.11.002

Roberts, G., Lewandowski, J., & Galantucci, B. (2015). How communication changes when we cannot mime the world: Experimental evidence for the effect of iconicity on combinatoriality. *Cognition*, *141*, 52–66. https://doi.org/10.1016/j.cognition.2015.04.001

Rubin, E. (1921). *Visuell Wahrgenommene Figuren*. Kopenhagen: Gyldendalske.

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

Russell, J. A. (2003). Core Affect and the Psychological Construction of Emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145

Russell, J. A. (2009). Emotion, core affect, and psychological construction. *Cognition and Emotion, 23*(7), 1259-1283.

Sanford, A. J., & Emmott, C. (2012). *Mind, brain and narrative*. Cambridge University Press.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology, 12*, 225–239. doi: 10.1037/h0070931.

Saussure, F. (1916/2011). *Course in general linguistics*. Columbia University Press.

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology*, *63*(11), 2251–2272. https://doi.org/10.1080/17470211003721642

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*(1–2), 227–256. https://doi.org/10.1016/S0167-6393(02)00084-5

Scherer, K. R., Clark-Polner, E., & Mortillaro, M. (2011). In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology*, *46*(6), 401–435.

Schiller, D., Freeman, J. B., Mitchell, J. P., Uleman, J. S., & Phelps, E. A. (2009). A neural mechanism of first impressions. *Nature Neuroscience*, *12*(4), 508–514. https://doi.org/10.1038/nn.2278

Schiller, F. (1981). *On the Naive and Sentimental in Literature*, translated and introduced by Helen Watanabe-O'Kelly, Manchester: Carcanet New Press.

Schimmack, U. (2005). Attentional interference effects of emotional pictures: Threat, negativity, or arousal? *Emotion*, *5*(1), 55–66. https://doi.org/10.1037/1528-3542.5.1.55

Schirmer, A., & Adolphs, R. (2017). Emotion Perception from Face,Voice, and Touch: Comparisons and Convergence. *Trends in Cognitive Sciences*, *21*(3), 216–228. https://doi.org/10.1016/j.tics.2017.01.001

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*(1), 24–30.

Schmidtke D.S., Conrad M. and Jacobs A.M. (2014). Phonological iconicity. *Frontiers in Psychology, 5*:80. doi: 10.3389/fpsyg.2014.00080.

Schmidtke, D. S., & Aryani, A. (2015). Bridges from affect to language Comment on

"The quartet theory of human emotions: An integrative and neurofunctional model" by S. Koelsch et al. *Physics of Life Reviews*, *13*, 83–84.

Schmidtke, D. S., Schröder, T., Jacobs, A. M., & Conrad, M. (2014). ANGST: affective norms for German sentiment terms, derived from the affective norms for English words. *Behavior Research Methods*. https://doi.org/10.3758/s13428-013-0426-y

Schröder, M., and Trouvain, J. (2003). The German text-to-speech synthesis system MARY: a tool for research, development and teaching. *International Journal of Speech Technology*, *6*, 365–377. doi: 10.1023/A:1025708916924

Schroeder, Mr. (1968). Reference signal for signal quality studies. *The Journal of the Acoustical Society of America*, *44*(6), 1735–1736.

Schröger, E., & Widmann, A. (1998). Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology*, *35*(6), 755–759. https://doi.org/10.1017/S0048577298980714

Schrott, R., & Jacobs, A. M. (2011). *Gehirn und Gedicht: Wie wir unsere Wirklichkeiten konstruieren*. Hanser.

Segal, E. (1995). Cognitive-phenomenological theory of fictional narrative. In Duchan, J., Bruder, G., and Hewitt, L., editors, *Deixis in narrative: a cognitive science perspective*, pages 61–78. Erlbaum, Hillsdale, NJ.

Seidenberg, M. S., & McClelland, J. L. (1989). A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review*. https://doi.org/10.1037/0033-295X.96.4.523

Semino, E. (2009). *Text Worlds. In 'Cognitive Poetics: Goals, Gains, and Gaps'*, Geert Brône and Jeroen Vandaele, (eds.), pp. 33-73, Berlin: Mouton de Gruyter, 2009

Semino, E. and Culpeper, J. (2002). *Cognitive Stylistics: Language and Cognition in Text Analysis*. Amsterdam and Philadelphia: John Benjamins.

Shapiro, M. (1998). Sound and meaning in Shakespeare's sonnets. *Language*, *74*(1), 81–103.

Shklovsky, V. (1990). *Theory of Prose*. Trans. Benjamin Sher. Elmwood Park: Dalkey Archive, 1990.

Sievers, E. (1912). *Rhythmisch-melodische studien: vorträge und aufsätze* (Vol. 5). C.

Winter.

Simner, J., Cuskley, C., & Kirby, S. (2010). What sound does that taste? Cross-modal mappings across gustation and audition. *Perception*. https://doi.org/10.1068/p6591

Simpson, P. (2004). *Stylistics, A Resource Book*. London: Routledge

Sokolova, M., Bobicev, V. (2009). Classification of Emotion Words in Russian and Romanian Languages. *Proceedings of RANLP- conference*, Borovets, Bulgaria, pp. 415-419.

Song, H., & Schwarz, N. (2009). If It ' s Difficult to Pronounce , It Must Be Risky. *Psychological Science*, *20*(2), 135–138. https://doi.org/10.1111/j.1467-9280.2009.02267.x

Sound Symbolism. (n.d.). In Wikipedia. Retrieved November 6, 2011, from en.wikipedia.org/wiki/Sound_symbolism

Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2007). Human brain activity time-locked to narrative event boundaries: Research article. *Psychological Science*. https://doi.org/10.1111/j.1467-9280.2007.01920.x

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, and Psychophysics*, *73*(4), 971–995. https://doi.org/10.3758/s13414-010-0073-7

Spence, C. (2012). Managing sensory expectations concerning products and brands: Capitalizing on the potential of sound and shape symbolism. *Journal of Consumer Psychology*, *22*(1), 37–54. https://doi.org/10.1016/j.jcps.2011.09.004

Stadthagen-Gonzalez, H., Imbault, C., Pérez Sánchez, M. A., & Brysbaert, M. (2017). Norms of valence and arousal for 14,031 Spanish words. *Behavior Research Methods*, *49*(1), 111–123. https://doi.org/10.3758/s13428-015-0700-2

Staub, A., & Clifton, C. (2006). Syntactic prediction in language comprehension: Evidence from either..or. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/0278-7393.32.2.425

Stenneken, P., Bastiaanse, R., Huber, W., & Jacobs, A. M. (2005). Syllable structure and sonority in language inventory and aphasic neologisms. *Brain and Language*, *95*(2), 280–292. https://doi.org/10.1016/j.bandl.2005.01.013

Stockwell, P. (2002). *Cognitive Poetics: An Introduction*. London: Routledge.

Taft, M. (1992). The body of the BOSS: Subsyllabic units in the processing of polysyllabic words. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 1004-1014.

Taub, S. F. (2001). *Language from the body: Iconicity and metaphor in American Sign Language*. Cambridge University Press.

Thompson, P. D., & Estes, Z. (2011). Sound symbolic naming of novel objects is a graded function. *Quarterly Journal of Experimental Psychology*, *64*(12), 2392–2404. https://doi.org/10.1080/17470218.2011.605898

Thompson, R. L., Vinson, D. P., Woll, B., & Vigliocco, G. (2012). The Road to Language Learning Is Iconic: Evidence From British Sign Language. *Psychological Science*, *23*(12), 1443–1448. https://doi.org/10.1177/0956797612459763

Thorndike, E. L. (1943). Man and his works.

Trost, W., Ethofer, T., Zentner, M., & Vuilleumier, P. (2012). Mapping aesthetic musical emotions in the brain. *Cerebral Cortex*, *22*(12), 2769–2783. https://doi.org/10.1093/cercor/bhr353

Trotsky, L. (1957). *Literature and Revolution*. New York: Russell and Russell, 1957.

Tsur, R. (1992a). *Toward a Theory of Cognitive Poetics*. Amsterdam: Elsevier.

Tsur, R. (1992b). *What Makes Sound Patterns Expressive?: The Poetic Mode of Speech Perception*. Durham: Duke University Press.

Tsur, R. (2003). *Aspects of Cognitive Poetics*. [Book Chapter] Full text available as: http://cogprints.org/3239/

Turner, M., & Lakoff, G. (1989). More than cool reason: A field guide to poetic metaphor. Chicago: University of Chicago Press.

Ullrich, S., Aryani, A., Kraxenberger, M., Jacobs, A. M., & Conrad, M. (2017). On the relation between the general affective meaning and the basic sublexical, lexical, and inter-lexical features of poetic texts-a case study using 57 Poems of H. M. Enzensberger. *Frontiers in Psychology*, *7*(JAN), 1–19. https://doi.org/10.3389/fpsyg.2016.02073

Ullrich, S., Kotz, S. A., Schmidtke, D. S., Aryani, A., & Conrad, M. (2016).

Phonological iconicity electrifies: An ERP study on affective sound-to-meaning correspondences in German. *Frontiers in Psychology*, *7*(AUG). https://doi.org/10.3389/fpsyg.2016.01200

van Bezooijen, R., & van Heuven, V. J. (1997). Assessment of synthesis systems. *Gibbon, D.; Moore, R.; Winski, R.(Ed.), Handbook of Standards and Resources for Spoken Language Systems*, 481–563.

Van Dijk , T. (1979). *Advice on Theoretical Poetics*, Poetics 8 (1979), 569-608

van Heuven, V. J., & Pols, L. C. (1993). *Analysis and synthesis of speech: strategic research towards high-quality text-to-speech generation* (Vol. 11). Walter de Gruyter.

Van Holt, N., Groeben, N. (2005). Das Konzept des Foregrounding in der modernen Textverarbeitungspsychologie. *Journal für Psychologie*, *13*, 311–332.

Van Peer, W. (1986). *Stylistics and Psychology: Investigations of Foregrounding*. London: Routledge Kegan & Paul.

Vega, F. C., Nosti, M. G., Gutiérrez, A. B., & Brysbaert, M. (2011). SUBTLEX-ESP: Spanish word frequencies based on film subtitles. *Psicológica: Revista de metodología y psicología experimental*, 32(2), 133-143.

Vigliocco, G., & Kita, S. (2006). Language-specific properties of the lexicon: Implications for learning and processing. *Language and Cognitive Processes*, *21*(7–8), 790–816. https://doi.org/10.1080/016909600824070

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*(02), 219–247. https://doi.org/10.1515/LANGCOG.2009.011

Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130292–20130292. https://doi.org/10.1098/rstb.2013.0292

Vinson, D., Thompson, R. L., Skinner, R., & Vigliocco, G. (2015). A faster path between meaning and form? Iconicity facilitates sign recognition and production in British Sign Language. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2015.03.002

Võ, M. L. H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research*

Methods, *41*(2), 534–538. https://doi.org/10.3758/BRM.41.2.534

Võ, M. L. H., Jacobs, A. M., & Conrad, M. (2006). Cross-validating the Berlin Affective Word List. *Behavior Research Methods*, *38*(4), 606–609. https://doi.org/10.3758/BF03193892

Võ, M., L.-H., Jacobs, A.M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A. & Hutzler, F. (2008). The Coupling of Emotion and Cognition in the Eye: Introducing the Pupil Old/New Effect. *Psychophysiology*, 45(1), 130-140

Vogt, N. (2018). Machine learning in neuroscience. *Nature Methods*, *15*(1), 33.

Wallace, S. (1982). Figure and ground: the interrelationships of linguistic categories. In P.J. Hopper (Ed.), *Tense-aspect: Between semantics and pragmatics* (pp. 201–223). Amsterdam: John Benjamins.

Wallis, J. (1699). *Grammar of the English language*. 5th ed. Oxford: L. Lichfield.

Walser, M. (1999) „Einer der auszog, das Fürchten zu verlernen“ in: Der Zorn altert, die Ironie ist unsterblich: Über Hans Magnus Enzensberger, ed R. Wieland (Frankfurt a. Main: Suhrkamp Taschenbuch), 18-22.

Wandell, B. A. (2011). The neurobiological basis of seeing words. *Annals of the New York Academy of Sciences*. https://doi.org/10.1111/j.1749-6632.2010.05954.x

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Warrington, E. K., & Shallice, T. (1980). Word-form dyslexia. *Brain*. https://doi.org/10.1093/brain/103.1.99

Watanabe-O'Kelly, H. (1981). Introduction. In: Schiller, F. *On the Naive and Sentimental in Literature*, Manchester: Carcanet New Press.

Watson, R. L. (2001). A comparison of some Southeast Asian ideophones with some African ideophones. *Typological Studies in Language*, *44*, 385–406.

Wells, J. C., (1997). SAMPA computer readable phonetic alphabet. In Gibbon, D., Moore, R. and Winski, R. (eds.), 1997. *Handbook of Standards and Resources for Spoken Language Systems*. Berlin and New York: Mouton de Gruyter. Part IV, section B.

Weninger, F., Eyben, F., Schuller, B. W., Mortillaro, M., & Scherer, K. R. (2013). On the acoustics of emotion in audio: What speech, music, and sound have in common. *Frontiers in Psychology*, *4*(MAY), 1–12. https://doi.org/10.3389/fpsyg.2013.00292

Westbury, C. (2005). Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain and Language, 93*, 10–19. doi:10.1016/

Whissell, C. (1989). The dictionary of affect in language. In Robert Plutchik and Henry Kellerman (Ed.), *Emotion: Theory, Research, and Experience* (pp. 113–131). New York: Academic Press.

Whissell, C. (1999). Phonosymbolism and the emotional nature of sounds: evidence of the preferential use of particular phonemes in texts of differing emotional tone. *Perceptual and Motor Skills, 89*(1), 19-48.

Whissell, C. (2000). Phonoemotional profiling: a description of the emotional flavour of English texts on the basis of the phonemes employed in them. *Perceptual and Motor Skills*, *91*(2), 617–648.

Whissell, C. (2002). Emotion conveyed by sound in the poetry of Alfred, Lord Tennyson. *Empirical Studies of the Arts*, *20*(2), 137-155.

Whissell, C. (2003). Reader's opinions of romantic poetry and consistent with emotional measures based on the Dictionary of Affect. *Perceptual and Motor Skills*, 96, 990 -992.

Whissell, C. (2009).Using the revised Dictionary of Affect in Language to quantify the emotional undertones of samples of natural language. *Psychological Reports*,105,1-13

Whissell, C. (2011). "To those who feel rather than to those who think:" Sound and emotion in Poe's poetry. *International Journal of English and Literature, 2*(6). 149-156.

Wiethoff, S., Wildgruber, D., Kreifelts, B., Becker, H., Herbert, C., Grodd, W., & Ethofer, T. (2008). Cerebral processing of emotional prosody-influence of acoustic parameters and arousal. *NeuroImage*, *39*(2), 885–893. https://doi.org/10.1016/j.neuroimage.2007.09.028

Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research*, *156*, 249–268.

Willems, R. M., & Jacobs, A. M. (2016). Caring About Dostoyevsky: The Untapped

Potential of Studying Literature. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2015.12.009

Winko, S. (2003). *Kodierte Gefühle: zu einer Poetik der Emotionen in lyrischen und poetologischen Texten um 1900* . Berlin: Erich Schmidt Verlag.

Wiseman, M., and Van Peer, W. (2003)Roman Jakobsons Konzept der Selbstreferenz aus der Perspektive der heutigen Kognitionswissenschaft. In H. Birus, S. Donat, and B. Meyer-Sickendiek (Eds.) *Roman Jakobsons Gedichtanalysen*, (pp. 277–306). Göttingen: Wallstein..

Wittfoth, M., Schröder, C., Schardt, D. M., Dengler, R., Heinze, H.-J., & Kotz, S. A. (2009). On emotional conflict: Interference resolution of happy and angry prosody reveals valence-specific effects. *Cerebral Cortex*, *20*(2), 383–392.

Wolf, W. (2005). The Lyric: Problems of Definition and a Proposal for Reconceptualisation. In E. Müller-Zettelmann, M. Rubik (Eds.) *Theory into Poetry: New Approaches to the Lyric,* (pp.21-56). Amsterdam: Rodopi. Wundt, W. (1896/ 2014). *Lectures on Human and Animal Psychology (Psychology Revivals)*. London: Routledge.

Wolfgang, I. (1976). Der Akt des Lesens. Theorie ästhetischer Wirkung. *München: W. Fink Verlag*.

Woodhead, Z. V. J., Barnes, G. R., Penny, W., Moran, R., Teki, S., Price, C. J., & Leff, A. P. (2014). Reading Front to Back: MEG Evidence for Early Feedback Effects During Word Recognition. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhs365

Wundt, W. (1900). Völkerpsychologie, vol. 1. *Die Sprache*, *2*.

Wundt, W. M. (1908). *Völkerpsychologie; eine untersuchung der entwicklungsgesetze von sprache, mythus und sitte: Bd. Die Kunst* (Vol. 3). W. Engelmann.

Wundt, W. M. (1922). *Die sprache* (Vol. 2). Alfred Kroner Verlag.

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*(5), 971–979.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*(2p2), 1.

Zajonc, R. B. (2001). Mere exposure: A gateway to the subliminal. *Current Directions* *in Psychological Science, 10*(6), 224-228.

Zajonc, R. B., Murphy, S. T., & Inglehart, M. (1989). Feeling and facial efference: implications of the vascular theory of emotion. *Psychological Review*, *96*(3), 395.

Zald, D. H., & Pardo, J. V. (2002). The neural correlates of aversive auditory stimulation. *NeuroImage*, *16*(3 I), 746–753. https://doi.org/10.1006/nimg.2002.1115

Ziegler, J. C., & Jacobs, A. M. (1995). Phonological information provides early sources of constraint in the processing of letter strings. *Journal of Memory and Language, 34*(5), 567–593.

Ziegler, J. C., Castel, C., Pech-Georgel, C., George, F., Alario, F. X., & Perry, C. (2008). Developmental dyslexia and the dual route model of reading: Simulating individual differences and subtypes. *Cognition*. https://doi.org/10.1016/j.cognition.2007.09.004

Ziegler, J. C., Ferrand, L., Jacobs, A. M., Rey, A., & Grainger, J. (2000). Visual and phonological codes in letter and word recognition: Evidence from incremental priming. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*. https://doi.org/10.1080/713755906

Ziegler, J. C., Jacobs, A. M. (1995). Phonological information provides early sources of constraint in the processing of letter strings. *Journal of Memory and Language, 34*(5), 567-593.

Ziegler, J. C., Montant, M., Briesemeister, B. B., Brink, T. T., Wicker, B., Ponz, A., … Braun, M. (2018). Do Words Stink? Neural Reuse as a Principle for Understanding Emotions in Reading. *Journal of Cognitive Neuroscience*, 1–10.

Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science, 12*, 379–384.

Ziegler, J. C., Tan, L. H., Perry, C., & Montant, M. (2000). Phonology matters: The phonological frequency effect in written Chinese. *Psychological Science, 11,* 234–238.

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. Psychology of Learning and Motivation, 44, 35–62.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation Models in Language Comprehension and Memory. Psychological Bulletin.

https://doi.org/10.1037/0033-2909.123.2.162

Zwaan, R.A. (1993): *Aspects of literary comprehension. A cognitive approach*. Amsterdam: John Benjamins.

Zwaan, R.A., Madden, C., and Stanfield, R. (2001). Time in narrative comprehension. P*sychology and sociology of literature*, 71–86.

Zwaan, R.A., Yaxley, R.H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychon Bull Rev*.,10(4):954-8.

List of Figures

words containing this specific phoneme. The PAP of each word was calculated as the average of all its PAVs. **B)** Words were synthesized and their extracted acoustic features were used in two multiple linear regression models as predictors for the PAP of arousal (right) and valence (left). The acoustic variables (11 in total) accounted for 27.9% and 23.7% of the variance in PAP_{aro} and PAP_{val} respectively (study 1).

Figure 6.2. Acoustic features of pseudowords (N=11) significantly predicted the ratings of their affective sound: 11.2% for valence (left) and 56.3% for arousal (right).

Figure 6.3. A) Acoustic profiles were constructed (using correlation cell plot) based on the strength and direction of correlations between the estimated effect of words' phonology on the evaluation of their affective meaning (i.e., Phonological Affective Potential: PAP), the two measures of words' affective sound (i.e., Affective Sound-Ratings: AS-R (study 2a), Affective Sound-Predicted: AS-P (study 2b), and ratings of words' affective meaning (i.e., Affective Meaning-Ratings: AM-R) on the one hand, and 11 acoustic variables on the other hand (left for valence, right for arousal). **B)** The correlation probabilities are shown in the table. Correlations not surviving Bonferroni correction for multiple comparisons are marked with "BF" (Bonferroni Failed).

Figure 6.4. A) The time course of sound intensity for the words "Gift /g ɪ f t/ *(gift)*" and "Stich /ʃ t ɪ ç/ *(stab)*" (top, yellow lines) compared to their counterparts "See /z e:/ (*lake*)" and "Lohn /l oː n/ (*wage*)" (bottom, red lines). Short vowels, plosives, and voiceless consonants (as in "Gift" and "Stich") possess smaller integrals of sound energy, whereas sustained high amplitude (see red lines) results in larger sound intensity. This relationship between phonetic features and sound intensity, together with the relationship between sound intensity and 'affective sound' of words, explains the harsh sound of words containing short vowels, plosives, and voiceless consonants. **B)** Spectral analysis shows that hissing sibilants in a word increase the sound's center of gravity (i.e., the magnitude-weighted mean of the frequencies present in the signal), which makes words including this category of phonemes sound harsh and negative (blue line Zwist /ts v ɪ s t/ *(strife)* vs. green line Lieb /l iː p/ *(kind)*).

Figure 7.1. Results of post-scan ratings were highly correlated with affective measures used for the fMRI-experiment. Left: lexical arousal ($r = .97$), Right: sublexical arousal ($r = .76$).

Figure 7.2. Words with a higher degree of lexical arousal (Lex H > Lex L) elicited stronger activation in a widespread network of medial and inferior frontal gyrus, as well as temporal pole, cuneus, precuneus, and posterior cingulate cortex. The reverse contrast (Lex L > Lex H) resulted in an enhanced BOLD signal in visual and somatosensory cortex ($p < 0.05$, FWE-corr.).

Figure 7.3. The main effect of *sublexical arousal* (i.e., words sounding high vs. low arousing) and the related pairwise comparisons were associated with an enhanced BOLD signal in bilateral posterior insula, superior temporal cortex (BA 22 extending to BA40), as well as supplementary and primary motor cortex (BA 6) ($p < 0.05$, FWE-corr.).

Figure 8.1. Words were organized in a 2 × 2 design with each of experimental factors (*lexical arousal* and *sublexical arousal*) manipulated in two distinct groups consisting of extreme levels of arousal (High = exciting, and Low = calming). The congruence vs. incongruence of *lexical arousal* (meaning) and *sublexical arousal* (sound) resulted in two groups of iconic vs. non-iconic words, respectively. Two example words (in German) from each category are given in each cell.

Figure 8.2. Congruent words (iconic) were classified more quickly (right) and more accurately (left) in the corresponding lexical group compared to incongruent words (non-iconic).

Figure 9.1. Word stimuli were organized in a 2x2 design: with each experimental factor (*lexical* and *sublexical arousal*) manipulated in two distinctive groups consisting of extreme levels of arousal (High=exciting, and Low=calming). The congruence vs. incongruence of *lexical* (meaning) and *sublexical arousal* (sound) results in two groups of iconic vs. non-iconic words, respectively.

Figure 9.2. Iconic words as defined by the congruence between *lexical* and *sublexical arousal* elicited BOLD signals in the left amygdala ($p<0.05$, FEW-corr). Pairwise comparisons showed increased activation in the same region for the contrast HH>HL, as well as LL>LH.

Figure 9.3. In the congruent condition (iconicity), left amygdala showed significant functional connectivity with activation in two seed regions: the left superior temporal gyrus (STG) and the left inferior frontal gyrus (IFG) representing the processing of sound and meaning of words, respectively.

Figure 10.1. Representation of the structure of a syllable (σ) using the example of the German word 'Gras'.

Figure 10.2. A random text is simulated as random binary sequences 000000100…. The symbol '1' appears with probability p and models a successful occurrence of a certain sub-syllabic unit in a text (in this example the phoneme ʊ), and the symbol '0' accounts for its unsuccessful occurrence with probability $1-p$.

Figure 10.3. Calculation of standard deviation for each set of pulling. Note that the relative frequency of each sub-syllabic unit is constant giving one function for each unit.

Figure 10.4. Binomial-based model for prediction of standard deviation using the examples of 'pf' and 'tS'.

Figure 10.5. The 3D-model for prediction of standard deviation of sublexical units as a function of text length and frequency.

Figure 10.6. Representation of workflow and modules employed in the tool

Figure 10.7. Hugo Ball's poem "Totenklage"

Figure 10.8. Diagram of phonemes in the poem "Totenklage" as outputted by the tool.

Figure 10.9. Diagram of "salient" syllabic onsets in the poem "Totenklage" as outputted by the tool.

Figure 11.1. The detailed procedure of analysis of poems based on Jakobson's model of language function, i.e., the communication between sender (the author, left side) and receiver (the reader, right side) through the message (poem, center). The ascription of a certain theme (affective meaning) to a text, understood as a form of meta-perception, is formed at both the supralexical (top) and the sublexical (bottom) level of language. Our measurement of the *basic affective tone* is purposed to capture the latter.

Figure 11.2. Calculation of sublexical measures of the *basic affective tone* for a given text. **A)** *sublexical affective values* (SAVs) of all syllabic units are calculated based on the average ratings of words containing a certain syllabic unit (see example of /kʁ/) **B)** A given text is phonemized using the G2P-software MARY and its salient syllabic units are subsequently extracted via a probabilistic model integrated in the "EMOPHON" **C)** The *basic affective tone* of the text is calculated based on the mean of SAVs for salient units. This mean value (Salient-SAV-Mean) is compared against an exhaustive distribution of random samples with matching numbers of units, to test for significance of deviations concerning SAVs – finally represented by *Salient-SAV-Sigma*.

Figure 11.3. A) Rating scores on five affective dimensions for three author-based categories **B)** The *basic affective tone* of arousal and valence as measured by *Salient-SAV-Sigma* for the author-based categories (top) as well as the reader-based categories of valence and arousal (bottom). The Y-axis represents the mean of *Salient-SAV-Sigmas* for each category.

Figure 12.1. The present work covers three out of four measurement classes for measuring literary experience, as suggested by Dixon and Bortolussi (2016).

Figure 12.2. Results of the empirical studies are summarized in three main categories concerning the three initial main questions (Chapter 1) of this dissertation.

Figure 12.3. A proposed extension of the Panksepp-Jakobson-Hypothesis as exemplified for perception of affective sound in language and poetry

Figure 12.4. A neurocognitive model of affective iconicity. A) The extension of a standard model of visual word processing to the effect of affective iconicity. B) The most critical brain regions associated with different representations. The color of the ovals denotes the theoretical representations in the model. Abbreviations: IFG, inferior frontal gyrus; SMG, supramarginal gyrus; STG, superior temporal gyrus; OP, occipital pole; AM, amygdala; VWFA, visual word from area.

Figure 12.5. The interaction of supralexical and sublexical processes leading to different emotional experiences as exemplified by Poe's "the raven".

Figure 12.6. An updated version of the NCPM that considers feedback loop from the resulted reading behavior back to the emotional experience during reading.

List of Tables

Table 10.2. Numeric representation of the salient phonemes in the poem “Totenklage”. Note that fractions appear since the value of confidence interval is calculated based on the mathematical model.

Table 10.3. Numeric representation of the sublexical salient units for all of 20 poems and newspaper texts. “Num” stands for the number of distinct phonological units being salient (the 1^{st} indicator) and “Sum” for the absolute sum of all segments positioned outside the confidence interval (the 2^{nd} indicator).

Table 11.1. Means and standard deviations of ratings on the affective dimensions arousal, valence, spitefulness, friendliness and sadness for poems from three different author-based categories (left). Pearson Product-Moment Correlation Coefficients (r) between rating scores of the five affective dimensions (right).

Table 11.2. Results of multiple-regression models for the prediction of ratings on five affective dimensions. Note that partial correlations are calculated based on the predictors that entered the stepwise regression model.

Table 12.1. A summary of different textual levels (single word vs. text), and presentation modalities (auditory vs. visual), as well as different methods and dependent variables used in each of the empirical studies. Dep. Var. = dependent variable, phon. = phonological, comp. = computational

Appendices

Study 1:

Phoneme	Number of Phonemes	PAV-Arousal	PAV-Valence
a	660	2.771966	-0.0592
ə	1533	2.694517	0.078447
pf	30	2.601649	0.087647
ts	243	2.89617	-0.17775
aː	418	2.748728	0.009111
b	417	2.681068	0.174536
au	164	2.643705	0.014397
d	280	2.725176	-0.00215
eː	316	2.732284	0.128438
ɛ	458	2.82648	-0.03215
f	493	2.795027	0.068395
g	374	2.811771	-0.00069
h	144	2.759515	0.171893
iː	397	2.815113	0.146776
ɪ	398	2.815126	0.11832
k	516	2.787631	0.093785
l	895	2.70632	0.033052
m	501	2.718058	0.039845
n	1308	2.756322	0.014857
ŋ	188	2.774024	0.180341
oː	368	2.773825	-0.03117
ɔ	223	2.805409	0.003907
p	339	2.812958	-0.01111
r	1324	2.829352	-0.01325
s	345	2.813455	-0.10704
ʃ	285	2.832619	-0.06539
t	1112	2.799762	0.015284
uː	176	2.807533	0.269463
ʊ	306	2.824526	-0.04343
v	224	2.771459	0.121828
ai	274	2.761602	0.133806
x	207	2.755794	0.022846
ɔy	55	3.053575	-0.32425
yː	38	2.809578	0.155573
ʏ	39	2.704151	0.427843
z	279	2.618919	0.214624

S1. The calculated values of *Phonological Affective Value (PAV)* for both valence and arousal for a list of 36 different phonemes in IPA.

Study 1:

Summary of Fit for PAP_{aro}:
RSquare: 0.282132
RSquare Adj: 0.27905
Root Mean Square Error: 0.021966
Mean of Response: 2.7685
Number of Observations: 2574

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2.73667	0.03980	68.75	<.0001
F0	0.00035	0.00009	3.63	0.0003
F1-Mean	0.00002	0.000003	7.80	<.0001
F2-Mean	0.00001	2.9e-6	3.93	<.0001
F3-Mean	0.00001	7.6e-6	1.67	0.0958
F1-Bandwidth	-0.00149	0.00038	-3.92	<.0001
F2-Bandwidth	-0.00012	0.00019	-0.64	0.5250
F3-Bandwidth	0.00005	5.8-6	8.70	<.0001
Intensity	0.00001	5.7e-6	1.92	0.0546
Intensity - SD	0.00004	6.1e-6	7.40	<.0001
Spectral CoG	-3.7e-6	4.2-6	-0.87	0.3825
Spectral SD	1.6e-6	2.2e-6	0.73	0.4638

Summary of Fit for PAP_{val}:
RSquare: 0.24034
RSquare Adj: 0.23708
Root Mean Square Error: 0.03267
Mean of Response: 0.035021
Number of Observations: 2574

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.10262	0.05913	1.74	0.0828
F0	-0.00004	0.00014	-0.33	0.7440
F1-Mean	-0.00006	5.5e-6	-11.75	<.0001
F2-Mean	0.00002	4.3e-6	6.20	<.0001
F3-Mean	7.1e-6	0.00001	0.62	0.5335
F1-Bandwidth	-0.00103	0.00056	-1.84	0.0666
F2-Bandwidth	0.00055	0.00029	1.86	0.0634
F3-Bandwidth	-0.00005	8.6-6	-6.71	<.0001
Intensity	3.8e-6	8.6e-6	0.45	0.6537
Intensity - SD	0.00003	9.1e-6	3.76	0.0002
Spectral CoG	-2.8e-5	6.3e-6	-4.42	<.0001
Spectral SD	-1.8e-6	3.4e-6	-0.55	0.5816

S2. Summary of regression models for PAP_{aro} (top) and PAP_{val} (bottom), predicted based on the 11 acoustic features of words

Study 1:

Summary of Fit for Ratings of Pseudowords / Valence:
RSquare: 0.12160
RSquare Adj: 0.11268
Root Mean Square Error: 0.36266
Mean of Response: 2.79591
Number of Observations: 1095

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-47.1406	8.07542	-5.83754	6.99E-09
F0	-0.00042	0.002185	-0.19442	0.845883
F1-Mean	-0.00046	0.000105	-4.34261	1.54E-05
F2-Mean	6.71E-05	8.18E-05	0.819816	0.412502
F3-Mean	3.33E-05	0.00012	0.278249	0.780874
F1 - Bandwidth	-0.00031	0.000162	-1.92997	0.053872
F2 - Bandwidth	-0.00012	9.94E-05	-1.19499	0.232351
F3 - Bandwidth	-9.1E-05	9E-05	-1.01467	0.310488
Intensity	0.71481	0.114618	6.236448	6.41E-10
Intensity - SD	0.000243	0.007009	0.034711	0.972316
Spectral CoG	-4.7E-05	5.33E-05	-0.88956	0.373898
Spectral SD	-3.4E-06	3.46E-05	-0.0994	0.920839

S3. Summary of the regression model for the valence ratings of pseudowords (affective sound), predicted based on 11 acoustic features

Study 1:

Summary of Fit for Ratings of Pseudowords / Arousal:
RSquare : 0.567404
RSquare Adj: 0.56301
Root Mean Square Error: 0.36168
Mean of Response: 2.95164
Number of Observations: 1095

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	130.5339	8.053664	16.20802	4.61E-53
F0	0.013187	0.002179	6.052399	1.96E-09
F1-Mean	0.000793	0.000105	7.554137	8.94E-14
F2-Mean	3.62E-05	8.16E-05	0.443408	0.65756
F3-Mean	-0.00042	0.000119	-3.51977	0.00045
F1 - Bandwidth	0.000773	0.000161	4.789613	1.9E-06
F2 - Bandwidth	-0.00025	9.92E-05	-2.56711	0.010388
F3 - Bandwidth	5.63E-05	8.98E-05	0.62661	0.531047
Intensity	-1.83045	0.114309	-16.0132	5.85E-52
Intensity - SD	0.020913	0.00699	2.991772	0.002836
Spectral CoG	7.81E-05	5.32E-05	1.46912	0.142091
Spectral SD	0.000214	3.45E-05	6.19966	8.03E-10

S4. Summary of the regression model for the arousal ratings of pseudowords (affective sound), predicted based on 11 acoustic features

Correlation Coefficients:

	PAP_{aro}	$AS\text{-}R_{aro}$	$AS\text{-}P_{aro}$	$AM\text{-}R_{aro}$		PAP_{val}	$AS\text{-}R_{val}$	$AS\text{-}P_{val}$	$AM\text{-}R_{val}$
F0	0.11	0.16	0.23	0.03		-0.02	-0.08	-0.04	0.02
F1-Mean	0.42	0.43	0.78	0.08		-0.38	-0.40	-0.85	-0.06
F2-Mean	0.24	0.30	0.47	0.04		0.026	-0.21	-0.31	0.02
F3-Mean	0.15	0.08	0.14	0.03		-0.03	-0.08	-0.19	0.01
Intensity	-0.35	-0.46	-0.80	-0.06		0.15	0.44	0.81	0.03
Intensity - SD	0.12	0.08	0.28	0.02		-0.004	-0.09	-0.19	0.01
F1-Bandwidth	0.42	0.43	0.71	0.08		-0.36	-0.38	-0.75	-0.05
F2-Bandwidth	0.31	0.27	0.45	0.04		-0.19	-0.26	-0.59	-0.01
F3-Bandwidth	0.29	0.20	0.23	0.09		-0.01	-0.12	-0.28	-0.01
Spectral CoG	0.25	0.37	0.64	0.05		-0.37	-0.30	-0.52	-0.07
Spectral SD	0.29	0.44	0.70	0.06		-0.33	-0.36	-0.56	-0.06

S5. Correlation Coefficients between PAP, AS-R, AS-P, AM-R and 11 acoustic variables.

Abbreviations: -aro=Arousal, -val= Valence, PAP= Phonological Affective Potential, AS-R= Affective Sound based on Word Rating (study 2a), AS-P= Affective Sound Predicted based on Acoustic Features (study 2b), AM-R= Ratings of words' Affective Meaning, i.e. original valence and arousal rating values in the database, BW= Bandwidth, SD= standard deviation

Study 5:

A1. List of Poems (Poet, Title):

1) Stefan George, Auf das Leben und Tod Maximins
2) Christian Morgenstern, Der Papagei
3) Ernst Moritz Arndt, Gottes Scherz
4) Gertrud Kolmar, Die Fremde
5) Heinrich Seidel, Hinter dem Kastanienbaum
6) Johann Gabriel Seidl, Wiederschein
7) Max Dauthendey, Der Himmel ward der Erde gleich
8) Conrad Ferdinand Meyer, In einer Sturmnacht
9) Ernst Moritz Arndt, Mut des Geistes
10) Gottfried Benn, Eingeengt
11) Johann Gabriel Seidl, Ausmarsch
12) Johann Wilhelm Ludwig Gleim, Abschied von Chloris
13) Karl Kraus, In diesem Land
14) Kurt Tucholsky, Vorfruehling
15) Ludwig Anzengruber, Die Herzenskuender
16) Oskar Loerke, Gestaltung
17) Rainer Maria Rilke, Das Fuellhorn
18) Stefan George, Zu meinen traeumen floh ich vor dem volke
19) Heinrich George, Es zuckt aus grauem wolkenzelt
20) Kurt Tucholsky, Deutscher Abend

Study 5:

A2. List of Newspapers' Articles (Newspapers' Name, Title Hyperlinked)

1) Bild: schwangere-stuttgarterin-setzt-hunde-und-katzen-aus
2) Braunschweiger Zeitung: forscher-praesentieren-meilenstein-beim-klonen
3) Frankfurter Allgemeine Zeitung:
bulgarien-wahlsieger-borissow-will-ergebnis-anfechten
4) FrankfurterRundschau: aerztefehler-keine-entschaedigung-trotz-behandlungsfehler
5) H-N-A: bombendrohung-kasseler-rathaus-arbeitsamt
6) Leipziger Volkszeitung
leipzig-faerbt-sich-schwarz-tausende-pilgern-zum-auftakt-des-wave-gotik-treffens
7) Münchner Merkur: fuerstenfeldbruck-stromausfall-stadtwerke-raetseln-ueber-ursache
8) Mitteldeutsche Zeitung: entlassung-von-wolff-boehmer-kritik-an-haseloff
9) Rheinische Post: zwei-angeklagte-wollen-aussagen
10) Spiegel: silicon-valley-reporter-google-glass-im-kurztest
11) Süddeutsche Zeitung,txt: us-praesident-obama-in-der-kritik-der-enttaeuscher
12) Abendzeitung:
malta-oder-san-marino-song-contest-giulia-siegel-blamiert-sich-auf-facebook
13) Berliner Zeitung: suff-pruegler-drischt-auf-busfahrer-ein
14) Göttinger Tageblatt: Warum-der-ICE-an-Goettingen-vorbeigerauscht-ist
15) Hannoversche Allgemeine Zeitung: Wohin-am-Pfingstwochenende
16) Maerkische Allgemeine:
Das-Landesverfassungsgericht-verhandelt-heute-die-Klagen-dreier-gut
17) Badische Zeitung: fuerchten-sich-suedbadens-buergermeister-vor-kitaklagen
18) Cellesche Zeitung: Wolf-reisst-Heidschnucke
19) Emder Zeitung: krise-ist-an-dramatik-nicht-zu-ueberbieten
20) Stimme (Heilbronn): Sattelzug-drohte-umzukippen

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt,

• dass ich die vorliegende Arbeit selbständig und ohne unerlaubte Hilfe verfasst habe,

• dass ich mich nicht bereits anderwärts um einen Doktorgrad beworben habe und keinen Doktorgrad in dem Promotionsfach Psychologie besitze,

• dass ich die zugrunde liegende Promotionsordnung vom 02.12.2008 kenne.

Berlin, den 19.06.2018 Arash Aryani