DE GRUYTER

Christiane Ulbrich, Alexander Werth, Richard Wiese (Eds.)

EMPIRICAL APPROACHES TO THE PHONOLOGICAL STRUCTURE OF WORDS

LA LINGUISTISCHE ARBEITEN

Empirical Approaches to the Phonological Structure of Words

Linguistische Arbeiten

Edited by
Klaus von Heusinger, Gereon Müller,
Ingo Plag, Beatrice Primus,
Elisabeth Stark and Richard Wiese

Volume 567

Empirical Approaches to the Phonological Structure of Words

Edited by
Christiane Ulbrich, Alexander Werth and Richard Wiese

DE GRUYTER

ISBN 978-3-11-054058-1
e-ISBN (PDF) 978-3-11-054289-9
e-ISBN (EPUB) 978-3-11-054064-2
ISSN 0344-6727

Library of Congress Cataloging-in-Publication Data
A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

Printing and binding: CPI books GmbH, Leck
♾ Printed on acid-free paper
Printed in Germany

www.degruyter.com

Contents

Christiane Ulbrich, Alexander Werth, and Richard Wiese
The word in phonology: questions and answers —— 1

Pia Bergmann
The phonological word in German – Insights from an acoustic-phonetic study of complex words —— 13

Dagmar Bronner, Nathanael Busch, Jürg Fleischer, and Erich Poppe
(Non-)separation of words in early medieval Irish and German manuscripts and the concept "word" —— 45

Javier Caro Reina
Word-profiling strategies in Central Catalan, Itunyoso Trique, and Turkish —— 71

Ulrike Domahs, Frank Domahs, and Christina Kauschke
The morphology-prosody interface in typically developing and language-impaired populations —— 95

Gerrit Kentner
Schwa optionality and the prosodic shape of words and phrases —— 121

Christiane Ulbrich and Richard Wiese
Phonotactic principles and exposure in second language processing —— 153

Alexander Werth, Marie Josephine Rocholl, Karen Henrich, Manuela Lanwermeyer, Hanni Th. Schnell, Ulrike Domahs, Joachim Herrgen, and Jürgen Erich Schmidt
The interaction of vowel quantity and tonal cues in cognitive processing: An MMN-study concerning dialectal and standard varieties —— 183

Natalie Boll-Avetisyan
The role of phonological structure in speech segmentation by infants and adults: a review and methodological considerations —— 213

Mathias Scharinger
Neural bases of phonological representations: Empirical approaches and methods — 241

Index — 273

Christiane Ulbrich, Alexander Werth, and Richard Wiese

The word in phonology: questions and answers

1 Introduction

In the history of linguistics, the notion of the word is a slippery one. Attempts at defining the “word” as a basic concept go back to the Greek grammarians such as Dionysios Thrax, but have never been completed or been satisfactory in all respects, as many definitions proposed are imprecise, contradictory or circular. Nevertheless, the word has remained a central concept in linguistic description and theory. One promising approach to the problem of defining the word in linguistics is to recognize that the word is not a single category, but one relevant to all major levels of linguistic description, and that it should thus be defined and studied with reference to each of these levels. This approach has at least the advantage of making the problem more manageable. In other words, we may postulate that there is a “word” in the sense of a morphological word, a syntactic word, a phonological word, a graphematic word, a lexeme (the set of word forms in a paradigm), and a semantic word. At the same time, the word is a unified concept, as it is a sign in the Saussurean sense, and therefore encompasses a level of form and a level of meaning.

All of these level-specific words mentioned above have been discussed, sometimes thoroughly, sometimes only cursorily; see, e.g., Julien (2006) on different notions of the word, Dixon and Aikhenvald (2002) on typological aspects, Gallmann (1999) on the graphematic word, Wurzel (2000) on the morphological word, and Hall and Kleinhenz (1999) on the phonological word. One initial crucial key factor relevant to for the distinction made between the word notions on the various linguistic levels derives from the fact that the different types of words are not in a one-to-one relationship. Consider so-called particle verbs such as *break up* in English or *aufhören* (‘to stop’) in German. Semantically, they are lexical units, but syntactically, they consist of two independent items occurring in different syntactic positions, as seen in *break it up* or *hörte damit auf* (‘stopped it’). Analogously, compounds in their entirety are seen as morphological or lexical units, while their component parts have been argued to form phonological words on their own (Wiese 1996: 72–74). The same is true at least for some morphologically derived words, which presumably have a different phonological status depending on the respective language (see below

DOI 10.1515/9783110542899-001

2.1). Again, words on different levels are non-isomorphic to each other. Such non-isomorphies are the basis of one line of argument for the separation of word notions on different, but systematically related, levels. A second line of argument for the separation of words stems from the fact that these word notions follow their own constraints and principles. We will pursue the latter in the following considering the word from a phonological point of view.

2 The word in phonology

2.1 Properties of the Phonological Word

The Phonological Word (PW) or alternatively the Prosodic Word can be defined, in first approximation, as a particular unit in the prosodic hierarchy. The PW is to be located within such a hierarchy between the lower unit of the foot and the higher unit of the Phonological Phrase (at least as far as the theoretical approach does not consider the Clitic Group as the unit immediately above the PW). This conception was most clearly postulated by Nespor and Vogel ([1986] 2007). They conceive this hierarchy as ranging from the syllable or the mora up to the Utterance. In contrast to most other prosodic units considered in the pertinent literature, the PW is defined with reference to morphological units by means of mapping rules: some types of morphemes form a PW of their own, while other types do not. The other such unit is the Phonological Phrase, which may be defined with reference to syntactic phrases.

The PW will turn out to be a useful, even necessary, category to the extent that it is needed as a domain that allows for an adequate explanation of various segmental, suprasegmental and phonotactic generalizations in the world's languages. For German, for example, Wiese (1996) and others argues that stems, prefixes and suffixes beginning with a consonant (e.g. *-heit* or *-lich*) correspond to an individual PW, while suffixes beginning with a vowel (e.g. *-ung* or *-ier*) or consisting of a single consonant (e.g. *-t*, *-s*) do not. Languages may vary in this respect; Booij (1984, 1999) argues that vowel harmony in Hungarian applies within the domain of the PW, and shows that *all* suffixes combine with the preceding stem in order to form the relevant PW domain for vowel harmony. Furthermore, the PW may provide the domain for the process of syllabification, a process related to phonotactic constraints such as the concatenation of consonants into word-initial and/or word-final clusters (cf. Raffelsiefen 2000; Wurzel 2000). The synchronization of segmental and suprasegmental units as well as temporal effects such as pausing, lengthening etc. is also applicable to the PW (cf. e.g. Prieto, Estebas-Vilaplana and del Mar Vanrell 2010 for F0-alignment in Catalan).

In addition, there are further processes identified in past research, including matters of stress assignment (one main stress per PW), assimilation, and dissimilation such as German nasal assimilation and degemination that appear to apply within the domain of the PW (cf. Wiese 1996). However, they have not satisfactorily been evidenced as diagnostic tools for the universal status of the PW (cf. Hall 1999: 17–19). More specifically, Dixon and Aikhenvald (2002: 13) state: “[i]t is clear that there is no single criterion which can serve to define a unit ‘phonological word’ in every language. Rather there is a range of types of criteria such that every language that has a unit ‘phonological word’ (which is probably every language in the world) utilises a selection of these.” Therefore, the discussion of the PW is not over, and there is clearly a need for extended empirical studies.

2.2 Boundaries and constituent parts of the Phonological Word

For the larger part of the history of phonology, the word, as a phonological object, was simply regarded as a linear sequence of phonemes. However, most researchers regardless of their respective framework will now agree that this is grossly over-simplified. The range of phonological categories below the PW encompasses at least the following: foot, syllable, syllable constituents (onset, rhyme, coda), and mora. The shape of phonological words is largely determined by universal and language-specific preferences in terms of these sub-constituents.

For example, the PW may be seen as containing preferentially a binary syllabic foot, or as minimally containing two morae (themselves forming either one strong syllable or two light syllables). Furthermore, the edges of PWs are often highlighted, either by a prominent stress on the final and/or initial syllable, by consonant clusters not possible within words, or by phonological rules applying at edges only.

The PW is used also to explain possible limitations on the size of words within languages. For example, there is a general tendency that PWs consist of no more than one lexical word (with the exception of Greek, see e.g. Nespor and Vogel [1986] 2007). Function words, by contrast, can only have the status of a PW when stressed. In unstressed position they can only be part of a PW (Peperkamp 1997: 16). Finally, at least in German PWs are minimally two-moraic, since they are to be located above the foot level within the prosodic hierarchy (minimality constraint).

In consequence, any thorough discussion of the PW will have to address questions of its word-internal constituents and boundaries. Some of the relevant issues are discussed in the present volume.

2.3 Relevant questions

The relevance of the PW for the explanation of phonological processes can be approached from various perspectives: at least those of speech processing, speech production, linguistic description and theory, and typological generalizations.

From the perspective of speech processing, there are three fundamental tasks to be solved by a hearer: segmentation of running speech, lemma identification, and integration, connected with the question whether the PW facilitates the decoding of running speech (Ulbrich and Werth 2017; Ulbrich and Wiese, this volume) or running text (cf. Bronner et al., this volume). Speech production, on the other hand, deals with the process of encoding phonological units into articulatory gestures (in the sense of Levelt 1992). A relevant question in this context is whether there is evidence for the PW in the production of speech, or more specifically if the realisation of segmental and suprasegmental units at PW-boundaries differs from those within the PW (cf. Wheeldon and Lahiri 1997; Bergmann, this volume). The status of the PW is relevant for linguistic description and theory to allow for integrated analyses at linguistic interfaces, i.e. within the field of morphophonology (cf. e.g. Peperkamp 1997; Domahs, Domahs, and Kauschke, this volume; Kentner, this volume). Finally, the evidence that the PW is a domain for the application of phonological processes provided in linguistic typology (cf. Auer 1993; Caro Reina, this volume) has led to the typological distinction between word languages and syllable languages (cf. e.g. Caro Reina and Szczepaniak 2014; Caro Reina, this volume).

3 Contributions to this volume

3.1 Empirical approaches

The major focus of the present volume is on the different ways in which the phonology of words can be studied empirically. Using the entire array of methods available to current research is one way of solving the puzzles existing around the notion of the PW. Major approaches documented in the different contributions of the present volume are the following:

Corpus studies: Bergmann carries out an acoustic-phonetic analysis, in order to test specific properties hypothesized for the PW in German. Bronner et al. present a historical and comparative perspective on the PW by measuring the origin and development of spaces between letters/words in a corpus of written texts from different languages and across a time span ranging from 700 to 1050.

Cross-linguistic analysis: Caro Reina contrasts stress assignment and phonotactic constraints as criteria for the PW in three languages. Such cross-linguistic comparisons also play a role in other contributions, for instance in Ulbrich and Wiese, and Werth et al.

Behavioral experiments: Domahs, Domahs, and Kauschke compare error patterns in complex words (participles) in children with SLI and typically developing children. Kentner presents the results of judgement and completion experiments with adult speakers on a range of rhythmical preferences suggested for the shape of the PW. Ulbrich and Wiese report the results of a reaction time study, using nonce words, to find evidence for principles proposed for consonant clusters on the right-hand edge of the PW.

EEG experiments: Domahs, Domahs, and Kauschke compare EEG responses to morphological violations in the processing of the PW by adults. Werth et al. also present the results of an EEG experiment on the processing of the PW on the bases of the interaction between vowel-quantity and tonal features in a Moselle-Franconian dialect compared to the standard variety of German. The neural representation of phonological features in the PW is addressed in Scharinger's survey, which also compares the values of EEG measurement to the fMRI paradigm.

Language acquisition: Language development is the focus of Boll-Avetisyan's literature survey. She discusses issues pertaining to the segmentation of the PW based on word-internal phonological structures in first and second language acquisition. As well as Scharinger's paper mentioned above, Boll-Avetisyan's paper presents a critical discussion of the empirical approaches and methods used in past and present research on the PW.

3.2 Topics represented

The chapter by **Pia Bergmann** investigates acoustic-phonetic properties of PW boundaries in complex words of German. In an attempt to distinguish boundaries of morphological and prosodic words, temporal measurements were obtained. On the basis of spontaneous speech data from different corpora, the duration of vowels and consonants in suffix-initial position was used to investigate if segmental duration depends on the morphological or prosodic structure of a PW (e.g. *händ.l+er* vs. *post.+ler* vs. *schwind.l+ig*). The factors 'complexity', 'frequency of occurrence' and 'semantic transparency' of the PW were included in the analysis. The statistical analysis revealed segment duration as a contributing factor for the constitution of prosodic word boundaries. However, some complex interactions were also found. Both frequency of occurrence as well as semantic transparency were factors explaining variance at PW boundaries. The results are interpreted in light of Turk's language redundancy thesis predicting prosodic structure to mirror redundancy factors including those investigated in Bergmann's analysis.

Dagmar Bronner, Nathanael Busch, Jürg Fleischer, and **Erich Poppe** investigate the implementation of the concept 'word' through the comparison of texts from Latin to Old High German, Old Saxon, and Old Irish vernacular manuscripts. At the end of the late-antique period and the beginning of the early medieval period, scribes began to graphically demarcate individual word-items. For lack of a generally agreed conceptualization of the 'word' at the time, no 'best practice' was to be employed by the scribes. Therefore, in the chapter a discussion of methodological issues impeding the identification of spaces in early medieval texts precedes the actual data analysis of prepositions or/and articles/demonstrative pronouns in combination with nouns in Old High German, Old Saxon, and Old Irish texts. The results reveal that phonological concepts such as stress groups were applied to separate graphic units in Old Irish. The separation of words (in a general sense) was found to predominate in Old High German and Old Saxon. However, units that are cross-linguistically found to be concatenated into clitics were also not spatially disjoined in the analyzed texts. Furthermore, a comparison of Old High German vernacular with Latin manuscripts by an individual scribe showed that the same strategies of demarcation were applied in both types of texts. Finally, an excursion discusses the word concept on the basis of grammatical teaching material for early medieval Irish.

A typological perspective is taken in **Javier Caro Reina**'s chapter differentiating three word languages Central Catalan, Itunyoso Trique, and Turkish. The three genetically unrelated languages are classified as stress-sensitive, distribution-sensitive, and harmonic word languages through the investigation of stress assignment and phoneme distribution. Caro Reina identifies language specific regularities in syllable structure and phonotactic restrictions sensitive to the prosodic unit of the PW. Central Catalan, classified as a stress-sensitive language, and Turkish, classified as a harmonic word language, show dependencies of vowel distribution to syllable structure and stress. Central Catalan vowels within the PW are characterized by diphthongization and lengthening as well as by shortening and reduction depending on stress, and it is the PW in Turkish which is characterized by vowel harmony. Additionally, reduced syllable complexity at word boundaries shows that Turkish syllable structure depends on phonotactic regularities. In Itunyoso Trique, by contrast, phonotactic regularities apply to the PW in that consonants are highly sensitive to within-word position.

The chapter by **Gerrit Kentner** reports three empirical studies concerning the relationship between optionality of schwa, rhythmic alternation, and prosodic parallelism in PWs and prosodic phrases. In a first study, two variants of the German adverb *gern/gerne* 'gladly' were presented within a sentential context and thereby creating stimulus sentences with a stress clash or a stress lapse. Participants were instructed to read out aloud the stimulus sentences. The results show that the graphemic representation of the adverb generally guides the participants' reading performance. Nevertheless, a small but significant effect for rhythmic environment was also found. The rhythmic structure of the environment on the left side of the target word was repeated in the realization of the adverb suggesting an effect of prosodic parallelism. The second experiment dealt with two syntactic alternatives in German possessive constructions (e.g. *der Knopf der Arbeitshose* 'the dungarees' button' vs. *der Knopf von der Arbeitshose* 'the button of the dungarees'). The two alternatives differ in rhythmic structure. In an online questionnaire, participants had to choose a monosyllabic or bisyllabic possessive following a monosyllabic or bisyllabic left environment in four target sentences. Again, the results indicate a preference for rhythm preservation. Prosodic parallelism did not appear to influence the participants' decision. Finally, the third study consisted of a corpus analysis of the aforementioned schwa-alternation of German adverbs, where stress-lapse, stress-clash, and instances of prosodic parallelism were queried. As in the previous two studies, rhythmic patterning appears to influence the distribution of the morphological alternatives. Evidence for the impact of prosodic parallelism was not found in this particular study.

The chapter by **Ulrike Domahs**, **Frank Domahs,** and **Christina Kauschke** deals with the interaction of prosody and morphology in the inflection and derivation of phonological words. Following a review of existing psycholinguistic evidence on the acquisition and processing of morpho-prosodic regularities, the authors discuss different phenomena of the German grammar before focussing on the prefixation of participles. Two studies are reported that compare morpho-prosodic abilities of subjects with a specific language development disorder to linguistically unaffected groups of subjects. The first study examines the prefixation in German participles in the production of speech-deficient children aged between 8 and 10 years by means of an elicitation task, and compares the results with an age-matched typically developing group and a younger control group. Neural correlates of processing violations of the morpho-prosodic participle prefixation rule in adults are considered in a second study. Electrophysiological responses are obtained from participants with a language impairment history in childhood are compared to a control group. The chapter brings together different aspects, i.e. the comparison of linguistic development in the performance of individuals with specific language development problems with linguistically unaffected individuals and the production and processing of grammatical operations at the morpho-prosodic intersection.

The chapter by **Alexander Werth, Marie Josephine Rocholl, Karen Henrich, Manuela Lanwermeyer, Hanni Schnell, Ulrike Domahs, Joachim Herrgen,** and **Jürgen Erich Schmidt** reports on the results of an EEG experiment investigating online processing of German PWs based on cross-varietal differing prosodic cues. Whilst many words in Middle-Franconian dialects (located in the western part of Germany) can be distinguished only by a combination of length and tone accent (e.g. [man^{1}] ‘basket’ vs. [manˑ^{2}] ‘man’), in Standard German prosodic cues do not suffice to distinguish meaning so that other acoustic cues, e.g. vowel or consonant quality have to be involved. Using a classic oddball paradigm, two groups of participants (Middle-Franconian dialect and Standard German speakers) were presented with a minimal pair ([ʃa̠ː^{2}l] ‘stale’ vs. [ʃa̠lˑ^{2}] ‘acoustic noise’) which have inverted lengths for the vowel and the lateral but both bear Tone Accent 2. Event-related potentials show that late mismatch negativity (MMN), resulting from pre-attentive processing, differs in amplitude and latency between the two participant groups: the MMN-component is generally shorter and its amplitude later in the dialect group. The authors suggest that these differences reflect differences in the phonological relevance of prosodic cues in the two varieties. Although both participant groups perceive Tone Accent 2 as a high tone, only the dialect group uses rules of tone-text-association within the minimal pair for lexical access.

The chapter by **Natalie Boll-Avetisyan** provides a comprehensive overview of experimental methods applicable in research of word-internal phonological structures that may facilitate or impede speech segmentation. In the first part of the chapter the so-called “segmentation problem” is introduced, followed by a discussion of some properties of phonological words. The author presents studies dealing with metrical cues such as trochaic rhythm, final stress and iambic rhythm as well as other suprasegmental cues such as syllable duration and pitch alternation previously shown to aid word segmentation and recognition in adult and child language acquisition. The role of phonotactic cues such as vowel harmony, vowel sequences and consonant clusters in adult native word segmentation and L2 word segmentation as well as probability cues for segmentation are further discussed in this chapter. Finally, the author points out advantages and disadvantages of existing experimental set-ups using natural and artificial languages, and provides suggestions for improving experimental methods in favour of an artificial language paradigm to be applied in experimental research on issues pertaining to word segmentation. The author also reflects on the current theoretical debate of an intricate interplay between universal and thereby innate features in contrast with language-specificity; both potentially shaping the phonological structure of words.

The thrust of **Mathias Scharinger**'s contribution is double-sided: on the one hand, the issue here is the appropriate level of abstraction to be assumed for the mental and neuronal representations of sounds in language. In this respect, the article argues the relevance of underlying forms and for the use of underspecification as the primary means for achieving an appropriate degree of abstraction. This type of abstraction per underspecification is crucially based on phonological features as the primary units. They may be either present or absent in lexical entries, and may only exert an influence if present. Furthermore, this contribution discusses the status of recent advances on the methodological level. Based on a comprehensive survey, the author argues that measurements of brain activity on the electrophysiological level (electroencephalography, EEG) or blood-metabolism level (functional magnetic resonance imaging, fMRI) provide valuable complementary tools for describing the place of phonology in the brain. The latter allows for a better coverage of spatial distribution whereas the former provides a higher temporal resolution of speech processing in the brain.

The chapter by **Christiane Ulbrich** and **Richard Wiese** reports on the results of a reaction time experiment. Applying an artificial language paradigm, the study explores how phonotactic knowledge of a first language affects the processing of complex syllable structures in a second language that may be either more or less constrained compared to the L1. In this study, the impact of a potentially universal phonotactic principle, i.e. that of sonority sequencing, is tested against actual occurrence as the crucial factor in usage-based models. Considering different proficiency levels in the second language, the impact of varying first language sonority requirements at the right edge of the PW and the influence of exposure on the processing of phonotactics in German as a second language is investigated. Russian learners are speakers of a language with a relatively complex syllable structure. The language is comparable to German in this respect, but crucially its clusters allow for frequent violations of the sonority sequencing principle. Chinese, on the other hand, is a language not allowing word final consonant clusters at all. Participants from both languages, Russian and Chinese, were found to be sensitive to the sonority sequencing principle regardless of their proficiency level in the target language German. Only the performance by advanced learners was influenced by the actual occurrence of a PW-final consonant cluster in the target language. Hence, the results of the study provide evidence not only for the interplay between universal and language specific phonotactic principles but also for their interaction with usage-based factors.

4 Concluding remarks

The present volume has its origins in a workshop with a title identical to that of the book. The conference took place in November 2014 in Marburg, and was hosted by the LOEWE Research Focus *Fundierung linguistischer Basiskategorien* (Exploring Fundamental Linguistic Categories) financed by the State of Hesse from 2012 to 2015. We thank the group of reviewers for their valuable work in improving on the contributions of this volume, all or our colleagues in the LOEWE Research Focus, and Marina Frank for her meticulous work on the production of the volume.

References

Auer, Peter. 1993. Is a rhythm-based typology possible? A study of the role of prosody in phonological typology (KontRI Working Paper 21). Konstanz: Universität Konstanz.

Booij, Geert. 1984. Neutral vowels and the autosegmental analysis of Hungarian vowel harmony. Linguistics 22. 629–641.

Booij, Geert. 1999. The role of the prosodic word in phonotactic generalizations. In Tracy A. Hall & Ursula Kleinhenz (eds.), Studies on the Phonological Word, 47–72. Philadelphia: John Benjamins.

Caro Reina, Javier & Renata Szczepaniak (eds.). 2014. Syllable and word languages. Berlin & Boston: De Gruyter.

Dixon, Robert M. W. & Alexandra Y. Aikhenvald (eds.). 2002. Word: A cross-linguistic typology. Cambridge: Cambridge University Press.

Gallmann, Peter. 1999. Wortbegriff und Nomen-Verb-Verbindungen. Zeitschrift für Sprachwissenschaft 18(2). 269–304.

Hall, Tracy Alan. 1999. The phonological word: a review. In Tracy A. Hall & Ursula Kleinhenz (eds.), Studies on the phonological word, 1–22. Philadelphia: John Benjamins.

Hall, Tracy A. & Ursula Kleinhenz (eds.). 1999. Studies on the Phonological Word. Philadelphia: John Benjamins.
Julien, Marit. 2006. Word. In Keith Brown (ed.), Encyclopedia of Language and Linguistics, vol. 13, 617–624. Amsterdam: Elsevier.
Levelt, Willem J. M. 1992. Accessing words in speech production: Stages, processes and representations. Cognition 42. 1–22.
Nespor, Marina & Irene Vogel. 2007 [1986]. Prosodic Phonology. Berlin: De Gruyter.
Peperkamp, Sharon. 1997. Prosodic words. The Hague: Academic press.
Prieto, Pilar, Eva Estebas-Vilaplana & Maria del Mar Vanrell. 2010. The relevance of prosodic structure in tonal articulation: Edge effects at the prosodic word level in Catalan and Spanish. Journal of phonetics 38. 687–705.
Raffelsiefen, Renate. 2000. Evidence for word-internal phonological words in German. In Rolf Thieroff, Matthias Tamrat, Nanna Fuhrhop, Oliver Teuber & Peter Eisenberg (eds.), Deutsche Grammatik in Theorie und Praxis, 43–56. Tübingen: Niemeyer.
Ulbrich, Christiane & Alexander Werth. 2017. Die Enklise von Präposition und Artikel in der Sprachverarbeitung. In Nanna Fuhrhop, Renata Szczepaniak & Karsten Schmidt (eds.), Sichtbare und hörbare Morphologie, 237–260. Berlin & Boston: De Gruyter.
Wheeldon, Linda R. & Aditi Lahiri. 1997. Prosodic units in speech production. Journal of Memory and Language 37. 356–381.
Wiese, Richard. 1996. The Phonology of German. Oxford: Clarendon Press.
Wurzel, Wolfgang Ullrich. 2000. Was ist ein Wort? In Rolf Thieroff, Matthias Tamrat, Nanna Fuhrhop, Oliver Teuber & Peter Eisenberg (eds.), Deutsche Grammatik in Theorie und Praxis, 29–42. Tübingen: Niemeyer.

Pia Bergmann

The phonological word in German – Insights from an acoustic-phonetic study of complex words

Abstract: A large body of research on the phonological word (ω) is focussed on its role for phonological processes like assimilation or degemination.[1] However, to date there are hardly any empirical studies that look at how phonological word structures are actually realized (cf. Hall 1999; Raffelsiefen 2000; Wiese 2000). The paper therefore sheds some light on the acoustic-phonetic realization of the phonological word boundary of complex words in German spontaneous speech. It investigates segment durations in the vicinity of morphological and/or prosodic boundaries in German derivations with V-initial suffixes and C-initial suffixes respectively. The analysis of roughly 600 items shows that phonological word structure indeed influences segment durations. Specifically, factors like word frequency expose an effect on prosodically complex words rather than on prosodically simple words.

Keywords: segment duration, suffix, German, spontaneous speech

1 Introduction

Since the late 1970s, the phonological word has been introduced into phonological theory to address persistent problems in morphophonology.[2] For example, Liberman and Prince (1977) assume the *mot* – an early expression for the phonological word – to play a major role in word stress assignment in compounds in English. In their account, phrases are supposed to carry main stress on the rightmost constituent, whereas compounds are usually stressed on the leftmost

1 I would like to thank the editors of the series and two anonymous reviewers for their helpful comments on earlier versions of this paper.

2 See Scheer (2011) for a comprehensive critical account of the history of approaches to the morphology-phonology interface. A recent overview over the phonological word can be found in Revithiadou (2011).

Pia Bergmann: Universität Duisburg Essen, Berliner Platz 6–8, D-45127 Essen, pia.bergmann@uni-due.de

DOI 10.1515/9783110542899-002

constituent so that *black bóard* (phrase) contrasts with *bláck board* (compound). However, when the rightmost constituent "branches", compounds are stressed on the rightmost constituent too. The relevant question then is what kind of constituent branching refers to. Liberman and Prince (1977) argue that it cannot be the syllable because compounds like *labor union* are stressed on the first constituent despite their branching syllable structure in *union*. Thus, the branching needs to refer to a different domain that is above the syllable but lower than the complex word. This domain is the phonological word (*mot*). Its main task is making the word-internal structure of the compound accessible for stress assignment. According to Giegerich (1985) compound stress in German behaves just the same so that words like *Weißweintrinker* ('drinker of white wine') bear main stress on the first constituent despite the second constituent (*trinker)*$_\omega$ ('drinker') being complex with respect to syllable structure as well as morphological structure. Words like *Weißweintrinkglas* ('white wine drinking glass'), however, carry main stress on the second constituent, where *trinkglas* ('drinking glass') consists of two phonological words: *((trink)*$_\omega$*(glas)*$_\omega$*)*$_\omega$.[3] Although more recent accounts of word stress have challenged the description of Liberman and Prince (1977) (see e.g. Bauer, Lieber, and Plag 2013), the idea of a prosodic constituent "phonological word" has steadily gained importance (see e.g. Nespor and Vogel 2007).

Hall (1999) proposes to subdivide the relevance of the phonological word (henceforth also pword, or *ω*) into three different areas: the regulation of phonological processes, the governing of phonotactic generalizations, and minimality constraints. In German, special attention has been paid to the first of these areas, i.e. the regulation of phonological processes. In this respect, segmental phonological processes like velar nasal assimilation and degemination are supposed to be governed by the phonological word meaning that both are obligatory within a phonological word but facultative or disallowed across a pword boundary as is exemplified by the following examples (cf. Wiese 2000: 68–69):

a. (U[ŋ]garn)$_\omega$
   'Hungary'
b. (u[n])$_\omega$(gern)$_\omega$ *or* (u[ŋ])$_\omega$(gern)$_\omega$
   'reluctantly'
c. (tan[ts]t)$_\omega$ *from* /tants/ + /st/
   'dance, 2. ps. sg.'

3 It will be assumed in this paper that phonological word structure is recursive (cf. Hall 1999; Wiese 2000).

d. (Schrif[t])$_\omega$([t]um)$_\omega$
'literature'

In words like *Ungarn* 'Hungary' (see Example a.), velar nasal assimilation is obligatory, whereas words like *ungern* 'reluctantly' can be produced with either an assimilated velar nasal or with an unassimilated alveolar nasal (see Example b.). Since the segmental strings under a. and b. are identical, the difference in their phonological behavior with respect to the assimilation has to be traced back to their different prosodic shape. In contrast to *Ungarn* the word *ungern* is prosodically complex, i.e. it consists of more than one pword, so that a phonological word boundary intervenes between the segments /n/ and /g/. Similarly, degemination of the alveolar fricative has to take place within a pword (see Example c.), but it should not occur when a phonological word boundary separates the two identical segments (see the preservation of the alveolar stop in Example d.).

In addition to segmental processes like the ones just described, the phonological word has an impact at the suprasegmental level, particularly on the process of syllabification. This can best be seen when comparing the syllabification of vowel initial suffixes to consonant initial suffixes. Formations with vowel initial suffixes (e.g. *-isch*, *-ung* or *-er*) result in structures where resyllabification across the morphological boundary is mandatory, e.g. *kin.d+isch* 'childish', *Mei.n+ung* 'opinion' or *Trin.k+er* 'drinker/alcoholic' as opposed to impossible **kind.+isch*, **Mein.+ung* or **Trink.+er*. (Syllable boundaries are indicated by ".", morphological boundaries are referred to by "+"). In formations with consonant initial suffixes on the other hand, the syllable boundary coincides with the morphological boundary: *täg.+lich* 'daily' or *frag.+los* 'unquestionable' even though resyllabification to *tä.g+lich* or *fra.g+los* would be required due to *Onset Maximization,* a principle stating that within a certain domain all segments that are allowed to build a cluster on phonotactic grounds in a specific language are syllabified as the onset of a syllable (cf. Vennemann 1982). The explanation for the syllabification of the suffixes is once again seen in a different prosodic structure of the two types of suffixes. An intervening phonological word boundary blocks the resyllabification into the onset of the second syllable in formations with consonant initial suffixes. Vowel initial suffixes are integrated into the phonological word of the stem so that resyllabification is possible: *(kin.disch)*$_\omega$ vs. *(täg.)*$_\omega$*(lich)*$_\omega$ (cf. Wiese 2000: 65–66).

Taking this as a starting point for a deeper examination of the phonological word in German, two points of criticism may be raised. First, none of the accounts considers empirical evidence for the assumed structures in any systematic way. Neither the arguments for the phonological word based on word stress patterns, nor those based on assimilation, degemination or syllabification are ever tested on existing speech materials. Hence, most of the descriptions rest upon introspection in the first place. Second, influences on pword boundary strength that go beyond a reference to speech rate or speech style (cf. Wiese 2000: 65) are neglected for the most part. This aspect is intertwined with the more general question whether gradient boundary strength should be of any relevance to the description of pword structure. In the more traditional approaches that were mentioned so far, gradience below or beyond a minimality threshold for the pword is not considered to be of any wider theoretical interest. Newer, more phonetically oriented approaches in prosodic phonology however, incorporate the thought of gradience into their modelling of prosodic structure. Specifically, Aylett and Turk (2004) and Turk (2010) propose the *Smooth Signal Redundancy Hypothesis* that links back to Lindblom's H&H-theory (1990) and inversely relates acoustic salience to language redundancy:

> I claim that the acoustic redundancy, or relative salience, of lexical words can be manipulated by signalling their boundaries. [...] Prosodic constituency is proposed to implement the relationship between language redundancy and word boundary salience.
>
> (Turk 2010: 231)

From this perspective, gradience of boundary strength is not some theoretically uninteresting matter of speech style or speech rate but an integral part of prosodic structure that is systematically influenced by certain linguistic factors subsumed under the notion of language redundancy. Figuring out the constitutive factors of language redundancy however, is far from trivial. A factor that certainly has to be incorporated into such a model is frequency (cf. Turk 2010: 230, 243). In the literature, several different frequency measures are discussed that contribute to language redundancy in the sense that higher frequency increases the predictability of an element. This predictability can either be context-bound, when frequency measures like backward/forward transitional probabilities calculate the probability of the occurrence of an element in co-occurrence with other surrounding elements. It can also be context-free, when the frequency of occurrence of an element in a text/corpus is captured regardless of its context. Although matters are more complex – especially with respect to transitional probabilities (cf. Pluymaekers, Ernestus, and Baayen 2005) – many studies underline the assumption that an increase in frequency diminishes acoustic salience, i.e. higher frequency furthers the acoustic reduction of elements (cf. Bell et al. 2009; Bush 2001; Bybee 2001, 2002; Jurafsky et al. 2001; Schäfer 2014; Zimmerer, Scharinger, and Reetz 2011, 2014; see also Ernestus and

Warner 2011). Linking back to the second point of criticism raised above, this means that not only is it necessary to investigate the production of the phonological word empirically but that an empirical study as such should also take into account influencing factors like lexical frequency of the items under investigation.

The present paper aims to shed some light on this issue. It investigates segment durations in morphologically complex words in German containing the suffixes *-lich*, *-ler*, *-er*, *-ig* and *-isch* and argues for the relevance of phonological word structure to explain the durational variation. Thus, the above-mentioned division into C-initial suffixes on the one hand and V-initial suffixes on the other is put to an empirical test. Moreover, the study systematically includes additional influencing factors like token frequency. The structure of the paper is as follows: The next section gives further insights into the prosodic structure and the phonetic realization of C-initial vs. V-initial suffixes in German. After that, the materials and methods of the present study will be described. Section 4 deals with the results of the study. The paper concludes with a discussion and conclusions.

2 The prosody and phonetics of C-initial and V-initial suffixes in German

The distinction between C-initial and V-initial suffixes holds a central position in the theoretical discussion of pword structure in German (cf. Hall 1999; Löhken 1997; Raffelsiefen 2000; Wiese 2000). In this discussion, the status of the V-initial suffixes is much more uncontroversial than that of the C-initial suffixes. As has been introduced above, V-initial suffixes are denied pword status based on the diagnostics of syllabification. Other diagnostics like word stress also speak against allocating pword status to the V-initial suffixes,[4] so that there is large agreement about their prosodic status (cf. Hall 1999; Löhken 1997; Raffelsiefen 2000; Wiese 2000). Focussing on syllabification only, the C-initial suffixes on the contrary are admitted pword status (cf. Wiese 2000). However,

4 This is true for native as well as non-native vowel-initial suffixes. Despite their often different behavior with respect to word stress (many non-native vowel-initial suffixes attract primary stress, whereas native vowel-initial suffixes hardly ever do so, one exception being *-ei* in *Meckerei* 'gripe' etc.), both stress patterns are compatible with stress assignment in simple phonological words (cf. Raffelsiefen 2000; Wiese 2000).

since syllabification is not the only possible tool for diagnosing pword status, conflicting results may arise, when other diagnostics are considered. One obvious example is the German diminutive suffix *-chen* (e.g. *Hündchen* 'little dog'), where the consonantal onset conflicts with the nuclear schwa vowel, which violates requirements necessary with respect to the minimality of phonological words, i.e. that they must contain a full vowel (cf. Hall 1999). Consequently, the pword status of the suffix *-chen* varies with the diagnostic that is deemed more important by the author.

Another consonant-initial suffix that has aroused particular interest is the suffix *-lich*. Several diagnostics come into play in the evaluation of this suffix. The picture that emerges as an outcome, is correspondingly diverse. Hall (1998) and Löhken (1997) claim that *-lich* does not qualify for a phonological word on the basis of its missing (secondary) word stress, whereas Raffelsiefen (2000) and Wiese (2000) assign pword status to *-lich* due to syllabification and the position of [ɪ] in the vowel inventory of German. Nübling et al. (2013: 75), finally, bring in the idea that *-lich* is currently on its way to losing its pword status, which they substantiate by referring to an ongoing change in possible syllabifications of formations with *-lich* (cf. Auer 2002, see below). From a diachronic point of view, *-lich* is related to OHG *lîch* 'body', a then freely occurring lexeme that developed into a bound suffix in the course of time. Although this correspondence with a diachronically related free lexeme is taken as an argument pro pword status for e.g. the suffixes *-haft* or *-schaft* by Löhken (1997), she does not consider this aspect for *-lich* and retains her analysis that *-lich* does not constitute a pword of its own.

As this short overview makes clear, assigning pword status to certain linguistic elements is a less than straightforward matter. Moreover, even if the problem of conflicting diagnostics was solved, we would still be encountering the methodological weakness that empirical proof for the assumed structures is largely missing, as was mentioned above. The lack of empirical evidence seems even more detrimental in the light of the fact that some of the assumed diagnostics may turn out to be inadequate altogether. This issue will be elucidated below by referring to a production study on final devoicing in German (Auer 2002).

Another basic assumption of the traditional accounts that has to be called into question is the assumption that the pword status is ascribed to each affix per se. That is, word specific differences in the bound forms of the affix are not taken into account. Obviously, this assumption runs counter to observed effects of frequency. If lexical frequency has a systematic influence on boundary strength, then the same suffix may differ with respect to its pword status. This view has the advantage that it offers a link between synchronic variation and diachronic change, too. Processes referred to as “high frequency fusion” in the traditional literature (cf. Raffelsiefen 2005) can thus be grounded in synchronic processes of systematic boundary reduction.

Until recently, phonetic studies on the realization of pword boundaries in German hardly existed at all (but see Bergmann 2012, 2014, forthcoming). One earlier exception is Auer (2002), who investigates the production of German words with C-initial suffixes as opposed to words with V-initial suffixes in read speech. As has been introduced above, these suffixes are supposed to differ in their prosodic structure, which surfaces as differences in syllabification – hence, words like *täg+lich* being expected to syllabify as *tä[k].lich* and words like *nebl+ig* as *ne.[b]lig* (and not **ne[p].lig* or **ne[pl].ig*). As indicated by the phonetic symbols, the syllabification should correspond to the application or non-application of final devoicing depending on the syllable position in which the sound comes to stand (cf. Wiese 2000). It is exactly this expectation concerning final devoicing that is cast into doubt by the results of Auer (2002) though: Most of the productions are characterized by final devoicing irrespective of the suffix-type. Consequently, final devoicing does not seem to be suitable to give an indication of a difference in prosodic structure. This does not mean, however, that V-initial and C-initial suffixes would collapse into one indistinguishable category. The results indicate that the suffix-types differ in another measure, i.e. the acoustic-phonetic cluster duration at the pword boundary instead. This finding is well in line with insights from phonetically oriented prosodic phonology, where, analogously to final lengthening on the intonation phrase level, lengthening in the realm of a pword boundary has been detected (cf. Sugahara and Turk 2009; Turk 2010 for an overview).

Conflicting results, on the other hand, are presented by Pluymaekers et al. (2010). They investigate complex words with the suffix *-(ig)heid* in spontaneous speech while focussing on the duration of the /xh/-cluster. Here, words like *zuinig+heid* ‘thriftiness’, *vast+igheid* ‘security’ and *baz+ig+heid* ‘bossiness’ not only differ in their deduced prosodic word structure. Additionally, the cluster is characterized by a different information load, which is estimated by counting the number of words in the paradigm with which the cluster in the word of interest contrasts. Unlike Auer (2002), Pluymaekers et al. (2010) do not find a correlation between cluster duration and prosodic structure. Their analysis reveals that the informational load in the morphological paradigm is a better predictor of cluster durations than prosodic structure. That is, the higher the informational load of the /xh/ sound sequence in the morphological paradigm, the longer its duration. This result fits in quite well with the assumption formulated by Turk (2010) that language redundancy serves as leverage in acoustic salience. It does not corroborate the idea though that language redundancy is implemented via prosodic structure.

Finally, the differing results by Auer (2002) and Pluymaekers et al. (2010) may be caused by methodological differences to a large degree. For one, Auer investigates read speech, whereas Pluymaekers et al. (2010) study spontaneous speech. Furthermore, statistical analysis is much more advanced in Pluymaekers et al. (2010), and in contrast to Auer (2002), they treat duration as a continuous variable while Auer makes a categorical distinction between longer and shorter clusters. As a consequence, it has to be stated that the results of Auer's pilot study seem to be rather preliminary. Still, both studies indicate that duration is a promising measure in the investigation of systematic influences of prosodic and/or morphological structure. It is this measure that will be taken up in the present study as main indicator of prosodic structure.

3 Materials and methods

The study is based on 610 tokens of morphologically complex German words uttered in spontaneous speech. The investigated corpora are the "DFG-Dialektintonationskorpus" (Au 72/13-1, Au 72/13-3, Au 72/13-4), the BigBrother-Corpus (Season 1), and the CallHome-Corpus (cf. Canavan, Graff, and Zipperlen 1997), all available in the *[moca]* database hosted at the University of Freiburg (http://moca.phil2.uni-freiburg.de/). The "DFG-Dialektintonationskorpus" comprises interviews with elderly speakers from different regions in Germany. The BigBrother-Corpus and the CallHome-Corpus both contain naturally occurring spontaneous speech in informal settings in face-to-face interactions (BigBrother) or telephone conversations (CallHome). Speakers are aged between approximately 20 to 60 years and stem from different regional backgrounds. Since dialectal variation was not in the center of interest in the present study, and due to a rather uneven distribution of speakers across different dialect areas as well as dialect levels, this aspect was not systematically investigated but rather controlled in the statistical analysis by introducing speaker as a random variable (see below).

In *[moca]*, complex words with the C-initial suffixes -*lich* and -*ler* and words with the V-initial suffixes -*ig*, -*isch*, and -*er* were selected and subjected to acoustic-phonetic analysis in Praat (cf. Boersma and Weenink 2013). Differences in lexical or semantic class of the complex words were not taken into account. The stem of the words with vowel initial suffixes always ends in /l/, so that the items show maximal similarity with those formations with the suffix *-lich*, e.g. in *(schwindl+ig)*$_{\omega}$ ‘dizzy’as opposed to *(end)*$_{\omega}$*(+lich)*$_{\omega}$ ‘finally’.

Table 1 gives an overview over the different suffixes by means of one test item each and indicates their prosodic structure as well as the location of the morphological boundary (+) and the syllable boundary (.). The number in the third column refers to the number of tokens with the given suffix that entered into analysis. Please note that the suffix *-ler* is analysed as forming a simple phonological word with the stem despite being consonant initial. This decision is due to its schwa nucleus vowel. In the test items, bold face marks the segments whose duration went into analysis.

Tab. 1: Data overview: Suffixes with their token frequency in the corpora analysed

Test item		Pword	n = 610	Translation
hän**d.**	**l**+er	one	103	‘dealer’
pos**t.**+	**l**er	one	79	‘post office worker’
schwin**d.**	**l**+ig	one	86	‘dizzy’
e**ŋ.**	**l**+isch	one	120	‘english’
en**d.**+	**l**ich	two	222	‘finally’

The **dependent variables** are duration (in s) of the boundary spanning sound sequence C1C2 (e.g. /dl/ in *-händler*) as well as duration (s) of C1 and C2 separately (marked in bold print in the table). Absolute durations were chosen in accordance with earlier studies in the realm of morphophonology/morphophonetics (cf. e.g. Pluymaekers, Ernestus, and Baayen 2005; Pluymaekers et al. 2010; Sugahara and Turk 2009).[5] Relevant influences on segment duration like syllable structure, manner of articulation and speech rate were separately coded and controlled via statistical analysis (see below).

All items were manually segmented based on visual inspection of the spectrogram and the waveform in Praat. Since the only sound kept constant across all investigated items is the /l/ (followed by /ɪ/ or /ɐ/), the preceding segmental context varies considerably covering vowels (e.g. *ehrlich* ‘honestly’) as well as affricates, fricatives, stops and nasals. Depending on the segmental context,

5 This is not to deny that relative duration can serve as an indicator of prosodic and/or morphological structure, as has been shown by e.g. Bergmann (forthcoming) and Plag, Homann, and Kunter (2017).

segmentation can become rather difficult (cf. Turk, Nakai, and Sugahara 2006). This is especially true for /l/ in intervocalic position because of the formant-like structure of the /l/ itself and its vulnerability for coarticulation (cf. Neppert 1999: 226–230). In addition to formant structure, amplitude changes in the waveform were therefore used as landmarks for the segmentation of /l/ (but see Turk, Nakai, and Sugahara (2006: 5, 15) for a word of caution concerning the segmentability of [l]). Qualitative differences in the realization of /l/ (like the occurrence of frication or multiple bursts) were not taken into account in the present study.

Figure 1 gives an instance of the segmentation of the test item *verständlich* (ge6333_336854). Here, /l/ is preceded by a voiceless alveolar stop which is part of a consonant cluster ([nt]).[6] C1 ([t]) coincides with a phase of silence and a burst. The segmentation of the beginning of the lateral is oriented to an increase in energy and the onset of a formant like structure. Likewise, the beginning of the following vowel is segmented in accordance with the onset of F2 and an increase in energy.

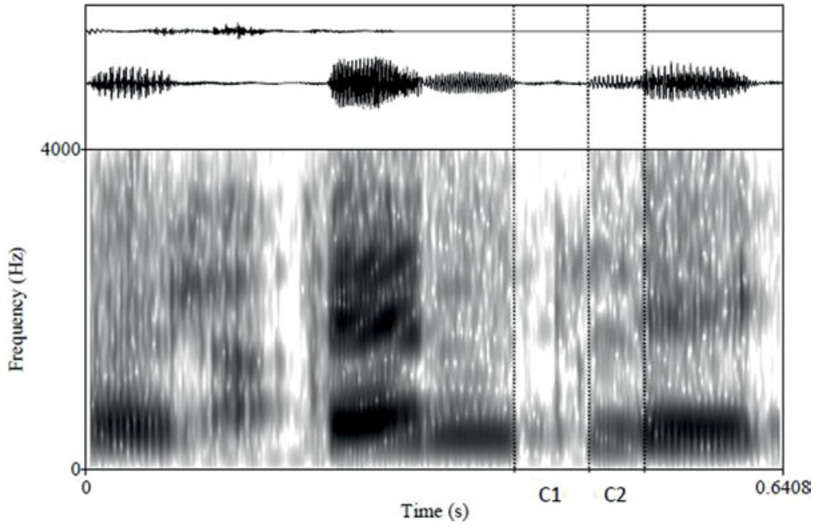


Fig. 1: Segmentation of the test item *verständlich* 'understandable' (CallHome, ge6333)

6 Segmentation of the stop generally included all phases (closure, burst and aspiration) if they were present. Voicing into the closure was not considered relevant for segmentation.

Items with deletions of C1, C2 or both were counted separately and were not integrated into the statistical analysis of the respective durations.

The **independent variables** are pword structure, lexical frequency and semantic transparency. **Pword structure** is as described in Table 1: Word formations with the suffix *-lich* correspond to a complex phonological word according to Wiese (2000) and Raffelsiefen (2000). Formations with *-er*, *-ler*, *-ig*, and *-isch* constitute a single pword together with the stem they are attached to.

Lexical frequency counts were obtained from COSMAS II (W – Archiv der geschriebenen Sprache) including all inflectional variants of the words. Excluded were words with extremely high frequency such as *eigentlich* 'actually' or *wahrscheinlich* 'probably', since these words are known to exhibit extreme forms of acoustical reduction which was not in the focus of interest in the present study (cf. Keune et al. 2005; Niebuhr and Kohler 2011). Ultimately, frequency covers a range from minimally 0 hits (e.g. *Filsbächler* 'a person who lives in Filsbach [a district in Mannheim]') to maximally 1.090.042 hits (for *Schüler* 'pupil'). For the statistical analyses, frequency was logarithmized (after adding 1 to each result) (cf. Baayen 2008: 31, 71–73).

As can already be seen in Table 1, the complex words differ with respect to **semantic transparency**. While the meaning of some of them is compositional like in *-händler* ('someone who deals'), others expose non-compositional meaning in at least some of their contexts of use, as e.g. *endlich* ('finally' or, transparent but less common, 'sth. that has an end'). As early as Giegerich (1985), it was assumed that semantic transparency could correlate with pword structure, which led us to include this factor into the analysis (see also Booij 1999 and Hay 2003 for more recent approaches to this issue). Considering that a lack of semantic transparency of the complex words renders the word's internal boundary less relevant, we assume that non-transparent items are characterized by decreasing duration of the segments adjacent to this boundary. Each individual item was considered with respect to its semantic transparency in its context leading to a categorical choice between 'transparent' and 'non-transparent'. In cases of doubt, online dictionaries (Wortschatz – Universität Leipzig, Digitales Wörterbuch der deutschen Sprache [DWDS]) were consulted by checking whether the stem occurred in paraphrases or synonyms for the word formation under investigation.

Since spontaneous speech is typically characterized by much variation that cannot be controlled beforehand, and since duration is susceptible to influencing factors like speech rate or number of syllables in the word, many **covariates** were additionally coded in order to control them during the data analysis by statistical means. These covariates are: manner of articulation in C1, rhyme structure of the syllable before C2, number of syllables per word, metrical structure of the word, the occurrence of a pitch accent on the word, the position of the word in the intonation phrase, and speech rate (measured as syllable/second according to the method of Fosler-Lussier and Morgan 1999).[7]

The **hypotheses** take up the assumptions concerning the relevance of prosodic structure as well as frequency of occurrence (see Section 1), as well as the effect of semantic transparency proposed by Giegerich (1985) (see above, this section):

(1) Complex pwords exhibit longer durations than simple pwords.
(2) Lower frequency items show longer durations than higher frequency items.
(3) Transparent items show longer durations than non-transparent items.

Additionally, for C1 an interaction between prosodic structure and frequency is expected. This expectation resides in the finding that domain final elements are specifically vulnerable to frequency effects (see e.g. Bybee 2001; Zimmerer, Scharinger, and Reetz 2011, 2014 on /t/-deletion). Since the C1 is domain final in complex pwords (like e.g. *(glück)*$_\omega$*(lich)*$_\omega$ 'happy') it should be more strongly influenced by frequency than the C1 in simple pwords as e.g. *(neblig)*$_\omega$ 'foggy', where it is not in domain final position. Accordingly, a fourth hypothesis was added:

(4) The reducing effect of frequency on C1 is stronger in complex pwords than in simple pwords.

Statistical data analysis was carried out by calculating mixed effects linear regression models (cf. Baayen 2008: 242–259) in R (R Core Team 2013). The packages used are *languageR* (cf. Baayen 2011), *lme4* (cf. Bates, Maechler, and Bolker 2013), *lmerTest* (Kuznetsova, Brockhoff, and Christensen 2016), *MASS* (cf. Venables and Ripley 2002), and *visreg* (Breheny and Burchett 2016). "Speaker" was included into the models as a random effect. Fixed effects were the abovementioned independent variables and the covariates as well as selected two-way-interactions. Models were fitted by stepwise including all predictors and by successively removing those that did not improve the model. Successive models were compared each by an ANOVA; further model criticism was based on qq-plots and explained variation (R^2). Differences of least squares means and pvalues were obtained for all fixed factors of the best-fit model using *lmerTest*. To avoid the danger of overfitting, care was taken that the number of tokens in

7 Syllable counts are based on canonical realizations of the utterance in this method. As a window for measurements, the intonation phrase was chosen.

the model was always at least 15 times as high as the number of predictors (cf. Baayen 2008: 195).

4 Results: Does prosodic constituency influence the duration of segments?

The section starts with an overview of the results for all durations (C1C2, C1, C2) (Section 4.1). It then focusses on the duration of C2 (which is /l/ in all test items) (Section 4.2). After the predictors introduced in Section 3 have been discussed, an alternative analysis will be presented that orients to the phonological surface structure of the test items rather than to the deduced phonological word structure. It includes the onset of the suffix (C-initial vs. V-initial, i.e. *-lich*, *-ler* vs. *-ig*, *-isch*, *-er*) and its vowel quality (full vowel vs. reduced vowel, i.e. *-lich*, *-ig*, *-isch* vs. *-ler*, *-er*) as predictors into the model, thus dispensing with the phonological word and testing, whether the segmental phonological structure per se may predict the durational outcome encountered in the test items (Section 4.3). On the basis of the results, it will be argued that phonological word structure indeed plays a major role for the durational realization of the segments in question and that it proves superior to the segmental make-up of the suffix as a relevant predictor for segmental duration.

4.1 The duration of the boundary adjacent segments

As described in Section 3, the durations of the sound sequence C1C2 as well as the durations of C1 and C2 separately were subjected to statistical analysis by fitting a mixed effects linear regression model for each of the durations. The models not only included the main predictors pword structure, frequency, and semantic transparency, but also several covariates like speech rate or manner of articulation of C1.

Table 2 summarizes the models by indicating which of the predictors, covariates and interactions reached a statistically significant t-value of 2.0 or above (cf. Baayen 2008: 248), or which improved the model as attested by model criticism and were therefore not removed from the model. The asterisk "*" indicates predictors with a t-value at or above 2.0; an asterisk in brackets indicates that the predictor reaches near significance with a t-value higher than 1.95. The last row gives the ratio of explained variation of each model (R^2). The main predictors are highlighted by bold face; "Sil/Sec" stands for "syllable/second", "SemTransparency" for "semantic transparency", and "IP" for "intonation phrase".

Tab. 2: Overview over the results for all durations[8]

Predictor	C1C2 (n = 328)	C1 (n = 315)	C2 (n = 545)
Pword structure	(*)	*	*
Frequency		*	
Semantic transparency	*		
Rhyme structure	*	*	*
C1 Manner of Articulation	*		
Metric structure	*		*
IP position	*		*
Speech rate (sil/sec)	*	*	*
Pword*Frequency		*	*
Pword*SemTransparency	*	*	*
Sil/Sec*SemTransparency	*		
Sil/Sec*Metric structure			*
Sil/Sec*Pword			
R^2 =	0.67	0.57	0.45

Without going into detail for every single influencing factor, a few things are worth pointing out: With respect to the main effects of the predictors of interest, at first sight only pword structure seems to exert a stable influence on all durational measures, as can be seen in the first row. However, a closer look reveals that frequency as well as semantic transparency enter into the models in interactions with other factors or with each other. This is of particular interest because an interaction between pword structure and frequency was expected for the duration of C1 (cf. Section 3). Indeed, the interaction confirms hypothesis (4) by ascribing a reducing effect of frequency to complex pwords as opposed to simple pwords. What was not part of the expectations spelled out in Section 3, though, is that the same interaction is valid for the duration of C2.

8 Complete best-fit models for the duration of C1C2 and the duration of C1 are given in an appendix to this article. The complete final model for C2 duration will be displayed and discussed in Section 4.2.

Figure 2 visualizes the interaction between frequency and pword structure for C1. The line in the left-hand column exemplifies the estimated duration of C1 in complex pwords, whereas the line in the right-hand column refers to duration of C1 in simple pwords. In both columns, the x-axis indicates token frequency (logarithmic), which is rising from left to right. (The interaction will be exemplified and discussed on the basis of the results for C2 as well, see below).

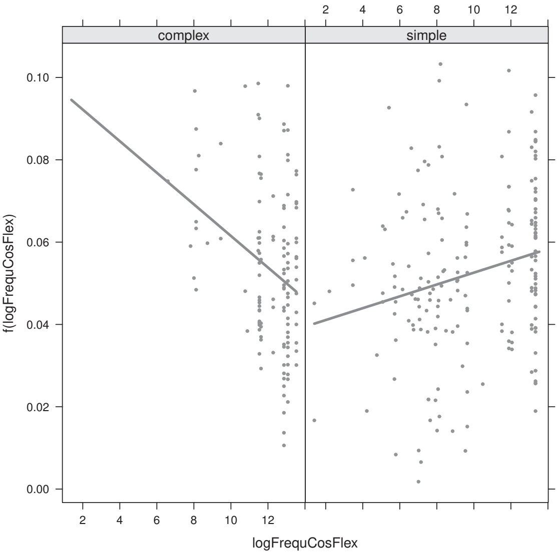


Fig. 2: Effect plot of frequency*pword structure (duration of C1 [s])

Figure 2 highlights the fact that frequency has a reducing effect on the duration of C1 in complex pwords (left column), whereas estimated duration of C1 increases with frequency for simple pwords (right column). Obviously, the behavior of simple pwords is not what was expected. One source of the problem may lie in the way we measured frequency: The frequency measures were always gathered for the complex word. Thus, if the high frequency item *Händler* 'dealer' occurred in words like *Blumenhändler* 'florist' or *Autohändler* 'car dealer', the frequency of the latter was recorded. This may well have influenced the results because most of these cases were restricted to formations with the suffix *-er* (i.e. *Händler* 'dealer', *Künstler* 'artist', *Sportler* 'athlete'). Given the well-established reductional effect of frequency in the literature, this can on the other hand be interpreted as a hint that the frequency of the immediate constituent may be of higher relevance than that of the complete word.

Semantic transparency, too, yields significant effects as a main predictor as well as in interactions with pword structure. This aspect will be touched upon when discussing the results for C2. All in all, the models show that the predictors indeed influence the durations of the segments which are adjacent to the prosodic and/or morphological boundary in the complex words under investigation. This can be seen as a first indication that boundary strength varies with pword structure, lexical frequency and semantic transparency.

Secondly, the covariates prove to be of major importance for the models (one exception being pitch accent which did not improve the predictive quality of the models). All durational measures are influenced by "rhyme structure" and by "speech rate". Thus, the more elements constitute the rhyme, the shorter the duration (cf. Klatt 1974). Likewise, increased speech rate leads to shorter durations. For C1C2, none of the covariates (except for pitch accent) could be dismissed from the model. These effects are neither surprising, nor are they at the centre of interest in the present study. Still, they underline the necessity to either control these aspects through careful design in experimental settings or to integrate them into statistical modelling if control is not an option, as is often the case in corpus studies of spontaneous speech.

4.2 Is there a difference between "stündlich" and "schwindlig"? Focussing on the /l/

In the next section, the results will be fleshed out in more detail on the basis of the duration of C2 because this is the model with the highest number of tokens available for the analysis. Recall that C2 always corresponds to the segment /l/. Thus, in the following, we compare the duration of /l/ in words like *(stünd)*$_{\omega}$*(lich)*$_{\omega}$ 'hourly' and *(schwindlig)*$_{\omega}$ 'dizzy'. The /l/ is pword-initial in complex phonological words but pword-internal in simple phonological words. It should be borne in mind that all complex pwords are constituted by instances of the suffix *-lich* in this study. The simple phonological words are comprised of instances of the suffixes *-ig*, *-isch*, *-ler*, and *-er*. The study may thus also be seen as an investigation of the suffix *-lich*, around which some controversies have arisen in the last decades, as was introduced in Section 2.

The analysis rests upon 545 tokens. Excluded were 40 items where the segment was deleted altogether. A few words shall be added on these **complete deletions of /l/** because their distribution may not be completely arbitrary and may probably lead to some interesting first insights. With respect to pword structure, 28 of the deletions are realized in complex phonological words (n = 190), whereas 12 deletions occur in simple phonological words (n = 355). That is, despite the presumably pword-initial position of the /l/ in formations with the suffix *-lich*, the segment does not seem to be particularly secure against extreme reductions. Moreover, with the exception of two instances (*verletzlich* 'vulnerable' and *verfänglich* 'captious'), all deletions in complex pwords occur in words with very high token frequency: *ursprünglich* 'originally', *zusätzlich* 'additionally', *verständlich* 'understandable', *gefährlich* 'dangerous', *ehrlich* 'honestly', and *kürzlich* 'recently'. This is not the case for simple pwords, where most of the deletions do not occur in high-frequency items: 9 deletions in moderately frequent words (*langweilig* 'boring', *künstlerisch* 'artistic', and *evangelisch* 'protestant/evangelical') are opposed to 3 deletions in highly frequent items (*Sportler* 'athlete' and *englisch* 'english').

Due to the few occurrences of deletions (n = 40), additional influencing factors like semantic transparency, speech rate or metrical structure were not systematically tested. The picture that emerges from the deletions may nonetheless serve as a preliminary basis for two possible conclusions: Either pword-initial position is not particularly safe against reductions, or pword status is weakened under the condition of high frequency. The fact that in simple pwords, frequency does not seem to play a decisive role for the occurrence of deletions, speaks in favor of the second conclusion rather. Considering that complex pwords under investigation include formations with the suffix *-lich* only, the results could also be viewed as a first indication of a loss of phonological substance in this particular suffix.

The idea of a gradual loss of boundary strength with increasing frequency in complex pwords will be further pursued when we now turn to the analysis of the durational reduction in C2. To give the reader a comprehensive picture of the interplay of all relevant predictors, the final mixed effects linear regression model will first be presented as a whole (Table 3). The discussion will then focus on the results for the main predictors, and the most relevant findings will be exemplified by effect plots (Figures 3–4).

In order to enhance readability, the predictors of main interest are highlighted by bold face. The interactions with the main predictors are displayed in italics. The last rows show the influence of the covariates. Significant influences are marked by an asterisk. A positive sign (+) indicates that the value given for each categorical predictor leads to an increase in acoustic duration, whereas a negative sign (–) means a decrease. The increase or decrease always refers to the baseline in the model (e.g. "PWord: one" for the predictor "PWord" or "Rhyme: V" for the predictor "Rhyme"). For the continuous variables (like frequency), the negative sign means that an increase (in frequency) predicts a decrease in duration.

Tab. 3: Best-fit mixed effects linear regression model for the duration (s) of C2

Random effects				
Groups	Name	Variance	Std. Dev.	
Speaker	(Intercept)	2.6947e–05	0.0051911	
Residual		2.0086e–04	0.0141723	
Number of obs: 545, groups: Speaker, 194				
Fixed effects				
	Estimate	**Std. Error**	**t value**	**sign**
(Intercept)	8.861e–02	7.471e–03	11.859	
PWord: two	**4.535e–02**	**1.277e–02**	**3.551**	*
Frequency	**–7.196e–05**	**3.374e–04**	**–0.213**	
SemTrans: no	**3.854e–03**	**2.494e–03**	**1.545**	
PWordtwo: Frequency	*–3.264e–03*	*1.050e–03*	*–3.107*	*
PWordtwo: SemTransno	*–1.133e–02*	*3.412e–03*	*–3.320*	*
Rhyme: VC	–1.299e–02	5.238e–03	–2.481	*
Rhyme: VCC	–1.813e–02	5.278e–03	–3.434	*
Rhyme: VCCC	–2.574e–02	5.815e–03	–4.427	*
Rhyme: VV	–4.199e–03	5.125e–03	–0.819	
Rhyme: VVC	–1.910e–02	5.454e–03	–3.502	*
Rhyme: VVCC	–1.179e–02	8.621e–03	–1.368	
Rhyme: VVCCC	–1.224e–02	1.161e–02	–1.053	
IP-Pos: medial	–9.175e–03	1.358e–03	–6.756	*
Sil/Sec	–3.889e–03	6.381e–04	–6.095	*
Metric: ww	–1.574e–02	7.185e–03	–2.191	*
Sil/Sec: Metricww	2.299e–03	1.134e–03	2.028	*

On the basis of this model ($R^2 = 0.45$), we can state that pword structure is a relevant predictor for the duration of C2. Complex pwords correspond to an increase in the segment's duration (t-value = 3.551). This main effect, however, cannot be interpreted without taking into account the relevant interactions. Pword structure significantly interacts with frequency (t-value = –3.107) as well as with semantic transparency (t-value = –3.32). First, the effect plot displays the interaction of pword structure with frequency (Figure 3).

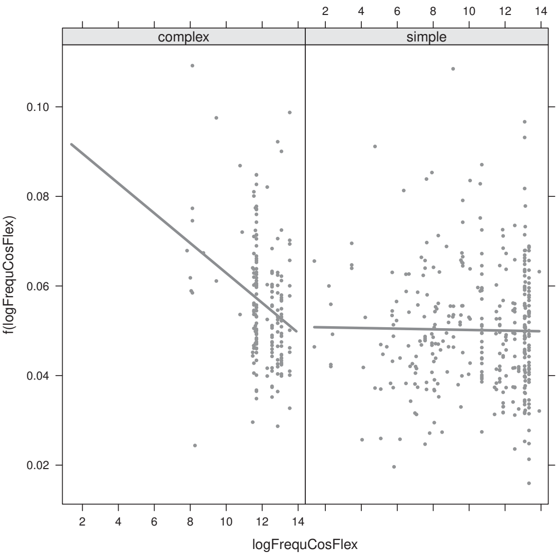


Fig. 3: Effect plot of frequency*pword structure (duration of C2 (s))

The effect of frequency in its interaction with the pword (Figure 3) is particularly interesting. As the model in Table 3 indicates, frequency does not qualify for a significant main effect, although the effect is in the expected direction (indicated by the negative sign, t-value = –0.213). However, despite the missing main effect on all items, frequency exerts a strong influence on the complex pwords, as is exemplified in Figure 3. Thus, while frequency does not seem to have any predictive power for the simple pwords,[9] the complex pwords are characterized by a significant durational decrease with increasing frequency, leading to nearly similar durations in the high frequency items. It can be concluded that simple and complex phonological words are characterized by different durational patterns, but these differences disappear in high frequency words.

Coming back to the preliminary thoughts that were formulated in connection with the deletions above, the patterns of durational reduction of /l/ in complex pwords gives us a clear indication that gradient weakening of the boundary is indeed taking place. There is thus evidence for the suffix *-lich* to gradually

9 In order to exclude the possibility that the result is due to the effect of the presumably problematic items with *-händler* as a second constituent in compounds, an alternative model without these elements (n = 22) was calculated which leads to similar results.

lose its pword boundary under the condition of high frequency, as was suggested by Nübling et al. (2013) (cf. Section 2). In addition, we find that hypothesis (4) formulated for the domain-final element C1 is also valid for C2, so that it has to be concluded that the position of the element with respect to a prosodic domain does not seem to play a significant role for its vulnerability to the reducing effects of frequency.

Interestingly, semantic transparency yields similar results with respect to its influence on complex vs. simple phonological words. The interaction between pword structure and semantic transparency (t-value = −3.32) is exemplified in Figure 4. Again, estimated durations in complex pwords are displayed in the left-hand column, and durations in simple words are given in the right-hand column. The x-axis refers to semantic transparency where "no" indicates non-transparent items and "yes" transparent items.

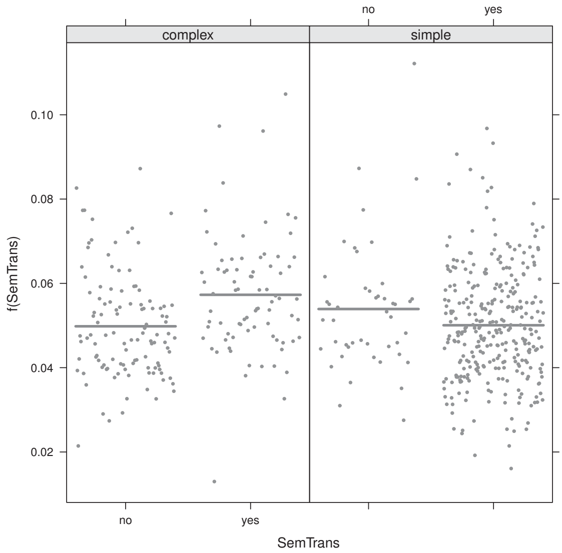


Fig. 4: Effect plot pword*semantic transparency (duration of C2 [s])

For the complex phonological words, non-transparent elements are predicted to have shorter /l/-durations than transparent elements (t-value = −3.10, p-value < 0.005). The simple phonological words however are not significantly influenced by semantic transparency (t-value = 1.55, p-value = 0.123) (cf. Figure 4; contrasts were tested for all fixed factors of the best-fit model using *lmerTest*).

The results for the complex phonological words thus confirm the hypothesis that semantic non-transparency leads to durational reduction. The simple pwords remain unaffected by this factor. One reason for this result may reside in issues of data distribution: In the group of the simple pwords, only 45 non-transparent items are opposed to 310 transparent items. The complex pwords on the other hand are more equally distributed with 110 non-transparent items and 80 transparent items. It is thus questionable, whether the result is reliable due to the few non-transparent tokens with simple pword structure. A deeper investigation of this issue on a more balanced database is therefore left to future research.

To sum up, the analysis corroborates the assumption that segment durations in the vicinity of a morphological and/or prosodic boundary are subject to systematic variation. Besides factors like speech rate or syllable structure, the prosodic structure of the word plays a major role for the segment's duration. It is intertwined in interactions with frequency on the one hand and with semantic transparency on the other. Both interactions show that it is only for the complex pwords that the hypotheses are confirmed. The more specific hypothesis that the reducing effect of frequency on C1 should be stronger in complex pwords than in simple pwords turned out to be true, too. However, since the same effect was found on the duration of C2, the expectation does not seem to be justified any longer because it rested on the final position of C1 in the prosodic domain.

Table 4 summarizes the hypotheses and the findings for the duration of C2. Hypotheses that were confirmed are marked by a check mark; a check mark in brackets refers to an ambiguous result, mostly indicating that the factor is not significant as a main effect.

Tab. 4: Overview over the hypotheses and the results for the duration of C2 (s)

Hypothesis	
(1) complex pwords > simple pword	✓
(2) lower frequency > higher frequency	(✓)
(3) semantic transparency > semantic non-transparency	(✓)
(4) frequency*pword for C1 with two pwords > two pword	(✓)

From this study, we can draw the conclusion that in acoustic-phonetic production in spontaneous speech, complex phonological words are treated differently from simple phonological words. They are characterized by longer segment durations at the boundary and a higher vulnerability to additional influencing factors like frequency and semantic transparency.

One legitimate objection, however, could be that the results just mirror the theoretically deduced categories that have been considered in the analysis. That is, collapsing the different suffixes with their specific phonological make-up into prosodic categories and feeding these into statistical analysis, may very well conceal relevant influences that do not refer to prosodic structure but to other more surface oriented aspects. To address this issue, an alternative model was fitted to the database that replaces the factor "pword" by the two surface-oriented factors "onset" (= "C-initial" or "V-initial") and "nucleus" (of suffix) (= "reduced vowel" or "full vowel"). It thereby takes up the phonological criteria that are deemed relevant for the assignment of pword structure and asks whether they can serve as predictors for the observed durational pattern. The results of this alternative model will be briefly discussed in the next section.

4.3 An alternative to pword structure?

The analysis again focusses on the duration of C2 (/l/). A mixed effects linear regression model was fitted to the same database as above (n = 545). While the factor "pword" grouped the suffixes *-ig*, *-isch*, *-er* and *-ler* under one heading as opposed to the suffix *-lich*, in this analysis, for the factor "onset" the suffixes *-ig*, *-isch* and *-er* are classified as vowel-initial, whereas *-ler* und *-lich* are consonant-initial. The factor "nucleus" comprises the suffixes *-ig*, *-isch* and *-lich* in the category "full vowel" and the suffixes *-er* and *-ler* in the category "reduced vowel".

The best-fit model yields a ratio of 46% explained variation, which is nearly identical to the 45% of the model including the pword. For the sake of completeness, the full model is displayed in Table 5. In contrast to the model with the factor "pword" (Table 3), manner of articulation of C1 (C1Art) is a significant covariate, whereas rhyme structure was removed during the fitting process; speech rate and IP-position remain their relevance. Frequency and semantic transparency interact with manner of articulation, which complicates the interpretation of their main effects a bit. Since the main interest of this section lies in the possible effects of the segmental phonological make-up of the suffixes, we will focus on the factor "onset", "nucleus", and the necessary interactions.

Tab. 5: Alternative model for the duration (s) of C2 (/l/)

Random effects				
Groups	Name	Variance	Std. Dev.	
Speaker	(Intercept)	1.7434e–05	0.0041755	
Residual		1.9203e–04	0.0138575	
Number of obs: 545, groups: Speaker, 194				

Fixed effects	**Estimate**	**Std. Error**	**t value**	**sign**
(Intercept)	0.0692299	0.0202958	3.411	
Onset: V-initial	**0.0004066**	**0.0015916**	**0.255**	
Nucleus: fullvowel	**–0.0119730**	**0.0081822**	**–1.463**	
Frequency	**–0.0034192**	**0.0011484**	**–2.978**	*
SemTrans: no	0.0407708	0.0119805	3.403	*
Nucleusfullvowel: Frequency	**0.0018885**	**0.0007984**	**2.365**	*
Frequency: SemTransno	–0.0032002	0.0009308	–3.438	*
C1Art: Nasal	0.0119426	0.0228645	0.522	
C1Art: Stop	–0.0073152	0.0211544	–0.346	
C1Art: Vowel	0.0207568	0.0209927	0.989	
Sil/Sec	0.0031341	0.0022105	1.418	
IP-Pos: medial	–0.0087119	0.0013136	–6.632	*
Sil/Sec: C1ArtNasal	–0.0056354	0.0024676	–2.284	*
Sil/Sec: C1ArtStop	–0.0059933	0.0023766	–2.522	*
Sil/Sec: C1ArtVowel	–0.0077838	0.0023459	–3.318	*
Frequency: C1ArtNasal	0.0011656	0.0012451	0.936	
Frequency: C1ArtStop	0.0031167	0.0011696	2.665	*
Frequency: C1ArtVowel	0.0025320	0.0010920	2.319	*
SemTransno: C1ArtNasal	–0.0055997	0.0092906	–0.603	
SemTransno: C1ArtStop	–0.0128954	0.0068694	–1.877	
SemTransno: C1ArtVowel	–0.0034697	0.0065234	–0.532	

Most interestingly, the factor “onset” neither yields a significant main effect, nor is it involved in any significant interactions. This shows that the segmental onset per se does not contribute to the explanation of the durational pattern thereby lending support to the theoretical claim that the suffix *-ler* should not be grouped with the suffix *-lich*. The factor “onset” is not able to prevail over the factor “nucleus”, which figures as a significant predictor in interaction with frequency (t-value = 2.365). According to this interaction, however, the reduction effect of frequency is only valid in elements with a reduced vowel – i.e. formations with *-er* and *-ler* – as is demonstrated by the effect plot in Figure 5 (the abbreviations “red” and “full” refer to “reduced vowel” and “full vowel” respectively):

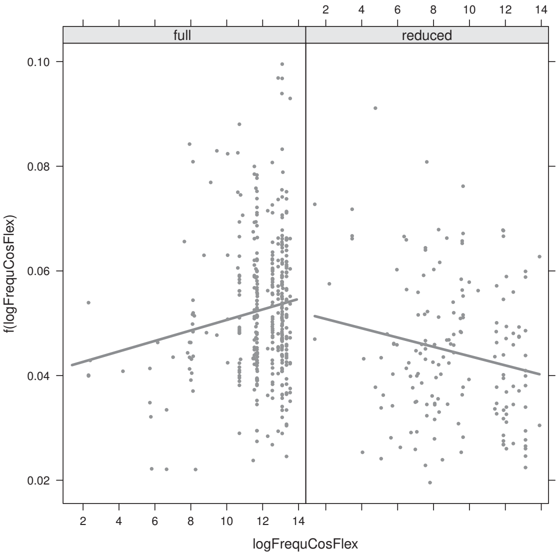


Fig. 5: Effect plot nucleus*frequency (duration of C2 [s])

Obviously, this interaction is not borne out by any of the expectations formulated earlier, and it is particularly peculiar in the light of the results that were gained for the data including the factor “pword”. Here, the complex pwords (= formations with the suffix *-lich*) were predicted to be influenced by frequency more strongly than the simple pwords. Since *-lich* contains a full vowel this is rather odd against the background of the present model, where elements with full vowel are predicted to increase C2 duration with increasing frequency. One source of this troublesome result could be that the suffixes were coded on a phonological basis adhering to the canonical realization with an [ɪ] for *-ig*, *-isch*, and *-lich* or [ɐ] for *-er* and *-ler*. It is thus possible that the division into elements with a full vowel on the one hand and elements with a reduced vowel on the other does not mirror the actual phonetic realizations of the items. For example, it could be that reduced instances of *-lich* under the condition of high frequency not only reduce the segmental duration at the boundary as has been shown in Section 4.2, but centralize the nucleus vowel, too. This would of course corroborate their gradient loss of pword status with increasing frequency. For the time being, this explanation belongs to the realm of speculation, though. What is more, it does not really explain why an increase in frequency should have a strengthening effect on C2 duration. To shed some light on this issue, the investigation of the phonetic realization of the nucleus vowel in these suffixes would be a worthwhile undertaking in future research.

As a summary of the presented model we can state that none of the segmental factors serves as a relevant predictor in terms of a main effect, while vowel nucleus surfaces in an interaction with frequency. Neither the onset of the suffix nor the vowel quality in the nucleus of the suffix seem to contribute clearly to an explanation of the durational variation of C2. It thus seems that the aspects of phonological surface structure in isolation do not play a major role for the durational reduction. This indirectly supports the claim of a prosodic structure where these aspects are combined – which then contributes to the explanation of the durational variation encountered, as was shown in Section 4.2. Hence, the introductory question if prosodic constituency has an influence on segment duration can be answered with a yes.

5 Discussion and conclusions

The present paper looked into the acoustic-phonetic realization of morphologically complex words in German spontaneous speech. Some of these words corresponded to complex phonological words and others to simple phonological words, according to phonological theory (cf. Hall 1999; Raffelsiefen 2000; Wiese 2000). The main aim of the paper lay in an empirical investigation of phonological word structure. It tested the influence of pword structure on durational realizations in the vicinity of the prosodic and/or morphological boundary in words like $(\textit{stünd})_\omega(\textit{+lich})_\omega$ 'hourly' and $(\textit{schwindl+ig})_\omega$ 'dizzy'.

The results demonstrate that the phonological word is in fact a relevant predictor for the investigated segmental durations: Segment durations are longer in morphologically complex words with an internal prosodic boundary. This finding thus corroborates the tentatively stated assumption by Auer (2002) that the durational pattern may be an indicator of prosodic structure in German. It contradicts Pluymaekers et al. (2010) though, who did not find an influence of prosodic structure on cluster duration in complex words in Dutch. Since

Pluymaekers et al. (2010) carried out a direct comparison between the impact of the prosodic structure of the complex words and the informational load of the cluster, however, the two studies are not entirely comparable. In the present study, morphological information did not enter into the analysis so that no conclusions can be drawn whether the pword would prevail under comparable conditions, or not. The comparison to the segmental phonological structure that was undertaken in the present study, led to the additional insight that neither the onset of the suffix nor the quality of its nuclear vowel plays a major role for predicting the segment's duration. This finding thus underlines the relevance of combining both aspects to form the phonological word, according to most accounts in phonological theory (cf. Hall 1999).

Given the fact that all complex pwords are instances of word formations with the suffix *-lich*, we also gained interesting insights with respect to the discussion around this suffix (cf. Section 2). First of all, we can state that boundary strength as measured by segment duration gradually decreases under the condition of high frequency. This can be seen as a synchronic evidence for the process sometimes called "High Frequency Fusion" (cf. Raffelsiefen 2000). While the description of this process rests upon sporadic evidence (e.g. English *cupboard* or *necklace*), we were able to demonstrate that the reducing effect of frequency can be depicted in synchronic data. It would be interesting to see if the effect found here is also valid for other suffixes with pword status in German, like *-heit*/*-keit* or *-tum*, or if it is specific for formations with *-lich*, which would support the assumption that *-lich* may be in the process of an ongoing change (cf. Nübling et al. 2013).

Moreover, the influence of frequency points to the fact that pword status should not be ascribed to suffixes (or likewise prefixes) per se because the status may very well depend on word-based information. Two confinements are in order here: Firstly, since the strong influence of frequency refers to one suffix only (namely *-lich*), more studies of other suffixes are necessary to substantiate this conclusion. Otherwise we would run the risk of basing the claim on just one suffix that may even be in the process of change, as was highlighted above. Secondly, this study focussed on boundary strength, but did not take into account other indicators of pword status, especially vowel quality in the suffix. Assuming a more holistic view of the phonological word, where more than one or two criteria serve to build a stronger or weaker phonological word, and investigating these phenomena on an empirical basis may be a fruitful direction to take in future research. Additionally, a systematic account of qualititative differences in the realization of /l/ and the preceding sounds – like a lateralization of a preceding burst or fricativization of /l/ – could complement the picture of boundary related phenomena. Another interesting strand of research could be to take a closer look at dialectal variation in the realization of prosodic structure – a factor that was controlled by statistical means in the present paper but not investigated in more detail.

To conclude this paper, we have been able to show that the constituent of the phonological word plays an important role for the acoustic-phonetic realization of complex words in German. It figures as an important domain for interactions with frequency and semantic transparency. This seems to confirm Turk's (2010) "Smooth Signal Redundancy Hypothesis" as it seems to be the case that factors of language redundancy are indeed implemented via prosodic structure. One challenge for future research lies in testing what factors may prove to be core aspects of this "language redundancy" – token frequency probably being only one of them. In a sense, many of the objections that came up in the course of this paper hint at the fact that we are just at the very beginning of testing and understanding the influences of prosodic and morphological structure on acoustic-phonetic realizations. Morphological structure per se does not seem to be a very stable influence (cf. Hanique and Ernestus 2012 for an overview), but certainly many more studies are necessary that delve into the details of different systematic influences in acoustic-phonetic realization of complex words in German and other languages.

References

Auer, Peter. 2002. Die sogenannte Auslautverhärtung in ne[b]lig vs. lie[p]lich – ein Phantom der deutschen Phonologie? In Michael Bommes, Christina Noack & Doris Tophinke (eds.), Sprache als Form: Festschrift für Utz Maas zum 60. Geburtstag, 74–86. Opladen: Westdeutscher Verlag.

Aylett, Matthew & Alice Turk. 2004. The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech. Language and Speech 47(1). 31–56.

Baayen, R. Harald. 2008. Analyzing Linguistic Data: A practical Introduction to Statistics Using R. Cambridge: Cambridge University Press.

Baayen, R. Harald. 2011. languageR: Data sets and functions with 'Analyzing Linguistic Data: A practical introduction to statistics'. R package version 1.4. http://CRAN.R-project.org/package=languageR (accessed 04 July 2016).

Bates, Douglas, Martin Maechler & Ben Bolker. 2013. lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-2. http://CRAN.R-project.org/package=lme4 (accessed 04 July 2016).

Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. The Oxford reference guide to English morphology. Oxford: Oxford University Press.

Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. Journal of Memory and Language 60(1). 92–111.

Bergmann, Pia. 2012. Articulatory reduction and assimilation in n#g sequences in complex words in German. In Philip Hoole, Lasse Bombien, Marianne Pouplier, Christine Mooshammer & Barbara Kühnert (eds.), Consonant Clusters and Structural Complexity, 311–343. Berlin: De Gruyter.

Bergmann, Pia. 2014. Reduction and deletion of glottal stops and geminates at pword-boundaries in German – Effects of word frequency and accentuation. In Javier Caro Reina & Renata Szczepaniak (eds.), Syllable and Word Languages, 251–278. Berlin: De Gruyter.

Bergmann, Pia. Forthcoming. Morphologisch komplexe Wörter im Deutschen. Prosodische Struktur und phonetische Realisierung (Studies in Laboratory Phonology 4). Berlin: Language Science Press. http://langsci-press.org/catalog/book/155 (accessed 12 December 2017).

Boersma, Paul & David Weenink. 2013. Praat: doing phonetics by computer [Computer program]. http://praat.org/ (accessed 04 July 2016).

Booij, Geert. 1999. The role of the prosodic word in phonotactic generalizations. In Tracy Alan Hall & Ursula Kleinhenz (eds.), Studies on the phonological word, 47–72. Amsterdam: John Benjamins.

Breheny, Patrick & Woodrow Burchett. 2016. Visualization of regression models using visreg. Version 2.3-0. http://myweb.uiowa.edu/pbreheny/publications/visreg.pdf (accessed 20 December 2016).

Bush, Nathan. 2001. Frequency effects and word-boundary palatalization in English. In Joan Bybee & Paul J. Hopper (eds.), Frequency and the emergence of linguistic structure, 255–280. Amsterdam: John Benjamins.

Bybee, Joan. 2001. Phonology and Language Use. Cambridge: Cambridge University Press.

Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. Language Variation and Change 14. 261–290.

Canavan, Alexandra, David Graff & George Zipperlen. 1997. CALLHOME German Speech LDC97S43. CD-ROM. Philadelphia: Linguistic Data Consortium.

Ernestus, Mirjam & Natasha Warner. 2011. An introduction to reduced pronunciation variants. Journal of Phonetics 39. 253–260.

Fosler-Lussier, Eric & Nelson Morgan. 1999. Effects of speaking rate and word frequency on pronunciations in conversational speech. Speech Communication 29. 137–158.

Giegerich, Heinz J. 1985. Metrical phonology and phonological structure: German and English. Cambridge: Cambridge University Press.

Hall, Tracy Alan. 1998. A note on secondary stress in German prosodic morphology. Linguistische Berichte 175. 414–424.

Hall, Tracy Alan. 1999. The Phonological Word: A Review. In Tracy Alan Hall & Ursula Kleinhenz (eds.), Studies on the Phonological Word. Amsterdam: John Benjamins, 1–22.

Hanique, Iris & Mirjam Ernestus. 2012. The role of morphology in acoustic reduction. Lingue e linguaggio 11(2). 147–164.

Hay, Jennifer. 2003. Causes and consequences of word structure. New York & London: Routledge.

Jurafsky, Daniel, Alan Bell, Michelle L. Gregory& William D. Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In Joan L. Bybee &

Paul J. Hopper (eds.), Frequency and the emergence of linguistic structure, 229–254. Amsterdam: John Benjamins.
Keune, Karen, Mirjam Ernestus, Roeland van Hout & R. Harald Baayen. 2005. Variation in Dutch – From Written “Mogelijk” to Spoken “Mok”. Corpus Linguistics and Linguistic Theory 1–2. 183–223.
Klatt, Dennis. 1974. The duration of [s] in English words. Journal of speech, language and hearing research 17. 51–63.
Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen Christensen. 2016. Tests in linear fixed effects models. R package version v2.0-33. http://CRAN.R-project.org/package=lmerTest (accessed 10 June 2017).
Liberman, Mark & Alan Prince. 1977. On Stress and Linguistic Rhythm. Linguistic Inquiry 8(2). 249–336.
Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. In William J. Hardcastle & Alain Marchal (eds.), Speech production and speech modelling, 403–439. Dordrecht: Kluwer.
Löhken, Silvia C. 1997. Deutsche Wortprosodie: Abschwächungs- und Tilgungsvorgänge. Tübingen: Stauffenburg.
Neppert, Joachim. 1999. Elemente einer akustischen Phonetik. Hamburg: Helmut Buske Verlag.
Nespor, Marina & Irene Vogel. 2007. Prosodic phonology. Berlin & New York: De Gruyter.
Niebuhr, Oliver & Klaus J. Kohler. 2011. Perception of phonetic detail in the identification of highly reduced words. Journal of Phonetics 39. 319–329.
Nübling, Damaris, Antje Dammel, Janet Duke & Renata Szczepaniak. 2013. Historische Sprachwissenschaft des Deutschen: Eine Einführung in die Prinzipien des Sprachwandels. 4. Auflage. Tübingen: Narr Verlag.
Plag, Ingo, Julia Homann & Gero Kunter. 2017. Homophony and morphology: The acoustics of word-final S in English. Journal of Linguistics 53. 181–216.
Pluymaekers, Mark, Mirjam Ernestus & R. Harald Baayen. 2005. Articulatory planning is continuous and sensitive to informational redundancy. Phonetica 62. 146–159.
Pluymaekers, Mark, Mirjam Ernestus, R. Harald Baayen & Geert Booij. 2010. Morphological effects on fine phonetic detail: The case of Dutch -igheid. In Cécile Fougeron, Barbara Kühnert, Mariapaola D’Imperio & Nathalie Vallée (eds.), Laboratory Phonology, 511–531. Berlin & New York: De Gruyter.
Raffelsiefen, Renate. 2000. Evidence for word-internal phonological words in German. In Rolf Thieroff, Matthias Tamrat, Nanna Fuhrhop & Oliver Teuber (eds.), Deutsche Grammatik in Theorie und Praxis, 43–56. Tübingen: Niemeyer.
Raffelsiefen, Renate. 2005. Paradigm uniformity effects versus boundary effects. In Laura Downing, Tracy Alan Hall & Renate Raffelsiefen (eds.), Paradigms in phonological theory, 211–262. Oxford: Oxford University Press.
R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://R-project.org/ (accessed 04 July 2016).
Revithiadou, Anthi. 2011. The phonological word. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), The Blackwell Companion to Phonology, 1204–1227. Malden, MA: Wiley-Blackwell.
Schäfer, Michael. 2014. Phonetic reduction of adverbs in Icelandic: On the role of frequency and other factors. PhD-thesis, Universität Freiburg.
Scheer, Tobias. 2011. A guide to morphosyntax-phonology interface theories. Berlin: De Gruyter.

Sugahara, Mariko & Alice Turk. 2009. Durational correlates of English sublexical constituent structure. Phonology 26. 477–524.
Turk, Alice. 2010. Does prosodic constituency signal relative predictability? A smooth signal redundancy hypothesis. Laboratory phonology 1(2). 227–262.
Turk, Alice, Satsuki Nakai & Mariko Sugahara. 2006. Acoustic segment durations in prosodic research: A practical guide. In Stefan Sudhoff, Denisa Lenertova, Roland Meyer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicole Richter & Johannes Schließer (eds.), Methods in empirical prosody research, 1–27. Berlin: De Gruyter.
Venables, W. N. & B. D. Ripley. 2002. Modern Applied Statistics with S. 4. Auflage. New York: Springer.
Vennemann, Theo. 1982. Zur Silbenstruktur der deutschen Standardsprache. In Theo Vennemann (ed.), Silben, Segmente, Akzente, 261–305. Tübingen: Max Niemeyer Verlag.
Wiese, Richard. 2000. The phonology of German. 2. Auflage. Oxford: Oxford University Press.
Zimmerer, Frank, Mathias Scharinger & Henning Reetz. 2011. When BEAT becomes HOUSE: Factors of word final /t/-deletion in German. Speech Communication 53. 941–954.
Zimmerer, Frank, Mathias Scharinger & Henning Reetz. 2014. Phonological and morphological constraints on German /t/-deletions. Journal of Phonetics 45. 64–75.

Online resources

COSMAS II (W – Archiv der geschriebenen Sprache). http://ids-mannheim. de/cosmas2/web-app/ (accessed 04 July 2016).
Digitales Wörterbuch der Deutschen Sprache (DWDS). http://dwds.de/ (accessed 04 July 2016).
Wortschatz – Universität Leipzig. http://wortschatz.uni-leipzig.de/ (accessed 04 July 2016).

Appendix

A. Best-fit mixed effects linear regression model for the duration (s) of C1C2.

Random effects				
Groups	Name	Variance	Std. Dev.	
Speaker	(Intercept)	0.00019723	0.014044	
Residual		0.00064110	0.025320	
Number of obs: 328, groups: Speaker, 143				

Fixed effects	**Estimate**	**Std. Error**	**t value**
(Intercept)	0.1934638	0.0169503	11.414
PWord: two	0.0509900	0.0255367	1.997
Frequency	−0.0002752	0.0009331	−0.295
Rhyme: VCC	−0.0199409	0.0086012	−2.318
Rhyme: VCCC	−0.0326890	0.0102906	−3.177
Rhyme: VVC	0.0073858	0.0090358	0.817
Rhyme: VVCC	−0.0151137	0.0150734	−1.003
Rhyme: VVCCC	−0.0003125	0.0208552	−0.015
C1ArtNasal	−0.0038200	0.0083650	−0.457
C1ArtStop	−0.0210046	0.0067807	−3.098
Sil/Sec	−0.0073059	0.0015413	−4.740
SemTrans: no	0.0652284	0.0234233	2.785
Metric: ww	−0.0152955	0.0056766	−2.694
IP-Pos: medial	−0.0129527	0.0034059	−3.803
SilWord	−0.0024308	0.0026145	−0.930
PWordtwo: Frequency	−0.0039196	0.0021701	−1.806
SilSec: SemTransno	−0.0080804	0.0036197	−2.232
PWordtwo: SemTransno	−0.0328294	0.0125324	−2.620

B. Best-fit mixed effects linear regression model for the duration (s) of C1.

Random effects			
Groups	Name	Variance	Std. Dev.
Speaker	(Intercept)	0.00011991	0.010951
Residual		0.00044628	0.021125
Number of obs: 315, groups: Speaker, 137			

Fixed effects	**Estimate**	**Std. Error**	**t value**
(Intercept)	0.1055053	0.0132176	7.982
PWord: two	0.0616831	0.0228830	2.696
Frequency	0.0014389	0.0006519	2.207
SemTrans: no	−0.0168644	0.0212922	−0.792
Rhyme: VCC	−0.0249924	0.0038810	−6.440
Rhyme: VCCC	−0.0359881	0.0054244	−6.634
Rhyme: VVC	0.0016060	0.0048192	0.333
Rhyme: VVCC	0.0014825	0.0107592	0.138
Rhyme: VVCCC	−0.0236355	0.0163480	−1.446
Sil/Sec	−0.0061643	0.0017146	−3.595
IP-Pos: medial	−0.0247364	0.0135808	−1.821
PWordtwo: Frequency	−0.0052759	0.0018886	−2.794
PWordtwo: SemTransno	−0.0335435	0.0146895	−2.284
Frequency: SemTransno	0.0033310	0.0026524	1.256
Sil/Sec:IP-Posmedial	0.0032419	0.0021554	1.504

Dagmar Bronner, Nathanael Busch, Jürg Fleischer, and Erich Poppe

(Non-)separation of words in early medieval Irish and German manuscripts and the concept "word"

Abstract: This chapter explores the implications of patterns of word division in selected Old High German, Old Saxon, and Old Irish manuscripts for the concept of the "phonological word". The replacement of *scriptio continua* in early medieval Latin and vernacular manuscripts by the introduction of spaces between linguistic elements ("words") has conventionally been associated with the writing habits of Irish scribes and scholars, who learnt Latin as a foreign language and who applied to their writing practice their grammatical knowledge about word classes. Patterns of word division in early vernacular manuscripts can therefore provide important indirect insights into the conceptualization of "words" in early stages of vernacular writing. Furthermore, little is known about the actual habits of Old High German, Old Saxon, and Old Irish scribes and the frequency and distribution of patterns of word division. The analysis of the textual data is complemented by a discussion of concepts of "word" that can be extrapolated from the grammatical teaching of early medieval Irish grammatical tracts.

Keywords: Old High German, Old Saxon, Old Irish, word separation, phonological word, early vernacular manuscripts

Dagmar Bronner: Johannes a Lasco Bibliothek, Kirchstraße 22, D-26721 Emden, bronner@jalb.de
Nathanael Busch: Universität Siegen, Hölderlinstr. 3, D-57068 Siegen, busch@germanistik.uni-siegen.de
Jürg Fleischer: Philipps-Universität Marburg, Pilgrimstein 16, D-35032 Marburg, jfleischer@staff.uni-marburg.de
Erich Poppe: Philipps-Universität Marburg, Wilhelm-Röpke-Str. 6E, D-35032 Marburg, poppe@staff.uni-marburg.de

DOI 10.1515/9783110542899-003

1 Introduction

One of the topics of a late-medieval Welsh "bardic grammar" — these are handbooks of grammar, rhetoric, and metrics for prospective poets — is the successful written transmission of a poem.[1] The author stresses that each word should be written separately, "because it would not be possible to understand the poem if it were written in one furrow" (Owen 2016: 199), that is, without separating words. In illustration, he writes out a quatrain in this format and comments: "as there is a fault with the writing, there would be a fault in the declaiming of the poem" (Owen 2016: 199). What he criticises here as a serious impediment to understanding is a specific form of writing, namely the *scriptio continua*, which was used by Latin scribes for their own language, but became slowly replaced by the separation of words in the early Middle Ages, for arguably the very same reason: to reduce the cognitive costs associated with processing.

With regard to the writing of Latin, there appears to be scholarly agreement that the change from *scriptio continua* to the graphic demarcation of individual words took place between the end of the late-antique period and the beginning of the early medieval period and that this process originated in the British Isles (Bischoff 1990: 173; Saenger 1997: 84–99). It probably originated in seventh-century Ireland and then spread to Anglo-Saxon Britain, that is, among speakers of both Celtic and West Germanic languages who had acquired Latin as a foreign language. Furthermore, Parkes (1991: 3–4) observes that Irish scribes followed different strategies when writing Latin and their own vernacular:

> When Irish scribes copied Latin texts they soon abandoned the *scriptio continua* which they had found in their exemplars. Instead they adopted as the basis for their scribal practices the morphological criteria which they had encountered in the analyses of the grammarians: they set out the parts of speech by introducing spaces between words. [... W]hen the Irish first began to copy texts in their native language, features of the spoken language are reflected in the way in which the texts have been set down on the page. [...] In the earliest surviving records of Old Irish prose [...] those words which are grouped round a single chief stress, and which have a close syntactical connexion with each other, have been copied as a single unit.

Parkes thus implies that early medieval Irish scribes had an intuitive concept of a fundamental category within their vernacular's linguistic system, a stress unit, which is different from the modern notion of the graphic or syntactic word.

1 For an edition and discussion of the text, which was composed in the third quarter of the fourteenth century, see Owen (2016).

His suggestion provides one motivation for our research on concepts of “word” as reflected in scribal practice during a comparatively early stage of the writing of a vernacular, the other being a lack of empirical quantitative studies of the actual performance of scribes. Some of the results from our research will be presented in this paper, which is organised as follows: We first discuss the methodological problem of identifying spaces in early medieval manuscripts and how they were dealt with when collecting the data (Section 2). We then present Old Irish (Section 3) and Old High German and Old Saxon data (Section 4), in both cases concentrating on combinations of prepositions or/and articles/demonstrative pronouns with nouns, with a digression on the conceptualisations of “word” in a medieval Irish grammatical tract (Section 3.3), before we come to some conclusions (Section 5).

2 Methodology: identifying “spaces”

In modern print, (non-)separation of words is a binary phenomenon: either two graphic units are separated by a space, or they are not. Graphic words are, by definition, entities that are separated from other graphic words by spaces. “Non-separation” is a deviation from this rule: two words that are usually separated by a space are written together and form a single (graphic) word. In case of “separation”, on the other hand, one unit usually realised as one graphic word is written as two (or more) graphic words. For our analyses we departed from this modern understanding. If circular argumentation is to be avoided, forms of separation and non-separation found in medieval manuscripts must not be used: We do not know whether the concept of the “(graphic) word” was known among early medieval scribes (and to what extent it was followed in scribal practice). Therefore, the modern understanding of the syntactic word constituted the basis for data collection and classification. In line with this notion, individual words, segments of words or clusters of words were classified by describing the spaces between the letters. Of course, this notion of “word” has its value only in terms of methodology: it is not to be expected that it was in any form “real” for the early medieval scribes who produced the manuscripts that are analysed here. However, it is possible to observe deviations from the modern usage, which, in turn, allows to contrast early medieval and modern practice.

The notion of “space” is independent of the notion of “word”. It designates a distance between letters that is not covered by ink; especially, it designates a relative distance. Crucially, even if there is no “space” between two letters, this does not necessarily mean that the two letters touch each other: Especially in non-cursive forms of writing a minimal space between two letters will very often be found. This is illustrated in Figure 1: In the first example, the letters *u* and *m* do not touch, but the distance between them is extremely small; it is actually not larger than the distance between the other letters. In this case, we therefore see no space between *u* and *m*. By contrast, in the second example the distance between *u* and *m* is far larger than the distance between all other letters; for this reason, we posit a space here. In other words, a "space" is defined as a distance between two letters that is wider than the average distance between letters (cf. Bronner and Busch 2015: 521).

Fig. 1: "Space" as a wider distance between letters

As already mentioned above, while in modern printing the decision whether there is a space or not is binary, for manuscripts it is scalar. No definition exists that will tell us when the distance between two letters is perceived as (and, thus, is) a "space". The decision whether there are "spaces" between letters, or whether letters of a text are written in a manner that might be called "aerated" (according to Saenger 1997: 32), is subjective and made by individual readers, as can even be seen by the differing assessments given by modern paleographers (cf. Bronner and Busch 2015: 528–529). The notion of "space" cannot be presupposed as an unambiguous given term of analysis.

It would be feasible to measure spaces between letters in manuscripts in terms of absolute distance. However, there would be two flaws: Such a procedure would be extremely time-consuming and, more importantly, the results obtained from different manuscripts could hardly be compared. Even in modern printing the distance between two letters depends on the size and the font of the characters. Therefore, absolute numbers are hardly telling; instead, a relative measurement is necessary.

Several parameters which had already been used in previous research (see Saenger 1997: 27) were discussed for the current analysis, e.g., the extent of the letter *o* or the distance between the two minims of *n* (cf. Bronner and Busch 2015: 522; Busch and Fleischer 2015: 567). The relative measure to indicate distances between letters was ultimately derived from the average line height, or, more precisely, from the vertical extension of letters without ascenders and descenders, as can be seen from Figure 2:

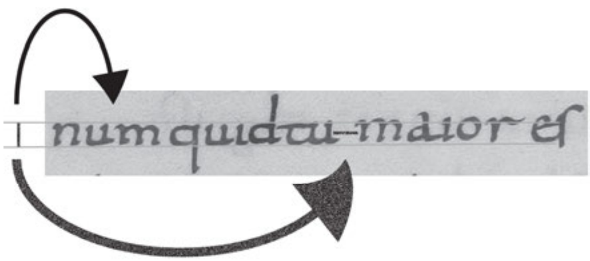


Fig. 2: Relative measure of the “average distance” between letters[2]

The line height, i.e. the average vertical extension of letters without ascenders and descenders such as e.g. *n*, *u*, or *m*, is put into a relation with the distance between letters; this then defines the “average distance” (for details on the methodology employed and our definition of “space” see Bronner and Busch 2015: 521–522 and Busch and Fleischer 2015: 565–569). We decided to follow this procedure because it allows for the comparison of different types of Insular and Carolingian minuscule and establishes a uniform standard for analysis. To be sure, “average distance” is no meaningful entity as such. It does not equal a “space” and has in no way any direct relevance for medieval scribal practice. Rather, it is an analytic tool to measure relative distances between letters.

Using “average distance” as a relative measure, the distances between letters of individual manuscript pages were recorded, employing values relative to the average distance (for instance, half of the average distance, one and a half of the average distance, etc.). For reasons of practicality, we introduced a notation system of fixed numerical values to represent or, rather, symbolize the

2 Figures 2 and 3 are based on St Gall, Stiftsbibliothek, Cod. Sang. 56 (http://www.e-codices.unifr.ch/de/list/one/csg/0056).

width of spaces in relation to the average distance. Figure 3 illustrates several distances between the letters *n* and *u*:

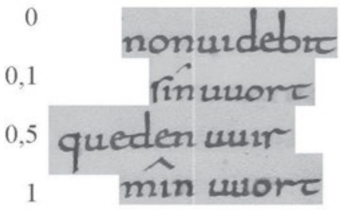


Fig. 3: Different distances between the letters *n* and *u*[3]

As the values on the left of Figure 3 indicate, the distance between *n* and *u* in *queden uuir*, for example, is approximately half as big as the distance in *mîn uuort*, while the distance between *sín* and *uuort* is very small and just distinguishable. Note that although we are using figures to indicate distances between letters, the resulting scale is ordinal, not cardinal. The figures are symbols and represent gradient values, i.e. our value 0.5 is used generally for distances which are clearly smaller than the average distance and still wide enough to be well recognisable as a "space" (and accordingly there are no values such as e.g. "0.3" or "0.7" in our scale). Thus, the values cannot be computed using their cardinal value.

Although in the end ten values were differentiated (most of them being larger than the "average distance"), for the present purposes it will suffice to apply a threefold distinction. First, the absence of a space (our value 0) is telling for our discussion. The actual size of a space, on the other hand, is only interesting insofar as it must be clearly discernible as a space between two letters. We consider this to be the case at a value of about half of the "average distance" (which corresponds to our value 0.5). This value, and all higher values, can therefore be conflated into one category "(clear) space". Finally, the third category designates minimal spaces which can neither be seen as clear cases of non-separation nor as unequivocal spaces (this is value 0.1 in Figure 3). It would not be justified to consider such cases either as instances of word separation or as instances of non-separation. The threefold distinction used in the remainder of

3 Due to a regrettable error the first and second line of Figure 3 in Busch and Fleischer (2015: 568) display mistakes. They are corrected here.

this paper is therefore made between "no space", "minimal space", and "(clear) space".

Necessarily, our point of departure was the medieval graphic "word". Stressed elements or the phonetic realisation generally were not taken into account. The units we refer to are letters and words. All distances between letters in selected passages were analysed for all (Old Irish, Old High German and Old Saxon, and Latin) records. After the collection the data were analysed statistically with the help of the programming language R.[4]

The data collection was exclusively based on examinations of manuscripts, facsimiles and high-quality digital images. Editions, even those that are otherwise very reliable, are not faithful with respect to the representation of separation and non-separation. This is almost to be expected since modern editions have to "translate" a scalar distinction into a binary one. Furthermore, for some Old High German records the editors attribute no value to the (non-)separation to be observed in the manuscripts and normalise according to modern standards (see Busch and Fleischer 2015: 569, with note 4), and this also holds for standardised editions of Old Irish texts.

3 (Non-)separation of words in Old Irish

Old Irish spans roughly the seventh to the ninth century, and is the earliest period of Irish from which a substantial contemporary corpus is extant, including mainly, but not exclusively, glosses of Latin texts. A neat summary of the current research status on word separation in Old Irish is given by Stifter (2010: 64), in the context of a description of Old Irish stress patterns:

> O[ld] Ir[ish] stress is [...] dynamic and fundamentally fixed on the first syllable of a word. [...] Articles, prepositions, conjunctions and various types of pronouns and pronominals are unstressed. Indeed, these can be regarded as pro- and enclitic to stressed words. Early Irish scribes used to write unstressed elements without separation from adjoining stressed words, a practice not followed by modern editors.[5]

4 Cf. e.g. Baayen (2014).

5 For the standard account of Old Irish scribal conventions regarding word separation see Thurneysen (1946: 24–25), the source for Parkes (1991: 3–5); for further references see Bronner and Busch (2015: 520, note 4) and Poppe (2016: 67–69). Importantly, some scholars have argued for the existence in the Celtic languages, and especially in Irish, of a word group built around a stressed word, variously termed "sentence word"/"Satzwort", "mot phonétique", or

Such descriptions are founded on scholars' informed intuitions about scribal practice, which are derived from their philological experience, but they are not backed by empirical research. This is where our project and its quantitative approach come in.

Before turning to some of the results from our analysis of the Old Irish corpus, one morphophonological feature of the language needs to be mentioned briefly, namely the existence of initial mutations. These are consonant changes in word-initial position triggered by a preceding word within phrases (see Stifter 2010: 65). Of the three types, lenition, nasalisation, and aspiration, only nasalisation is relevant for the present context because nasalisation is orthographically marked after a nasalising word on voiced stops and vowels, e.g., *a* ***n****gáe* 'their spear', *a* ***m****bó* 'their cow', *a* ***n****-enech* 'their honour', with nasalisation here triggered by the possessive pronoun third person plural *a*.

In the following, some results of our empirical quantitative analysis of the Old Irish corpus will be presented, which connect with grammaticographical reflections of medieval Irish scholars in interesting ways (see Section 3.3). The corpus consists of excerpts from the three main manuscripts containing Old Irish glosses, namely the Würzburg glosses on St Paul's epistles, the Milan glosses on a Latin commentary on the Psalms, and the St Gall glosses on Priscian's Latin grammar in Cod. Sang. 904. The marginalia and three poems transmitted in the same manuscript have also been analysed, as well as four incantations in Cod. Sang. 1395 and five poems in the Codex Sancti Pauli, a small collection of miscellaneous texts. There is little Old Irish consecutive prose; two narrative glosses in the Milan codex, the so-called *Additamenta* in the Book of Armagh, and a tract on the mass in the Stowe Missal are included in our corpus.

Tab. 1: Old Irish sources

Location	Date	Passage(s) analysed	Reference
Würzburg, UB, M. p. th. f. 12	ca. 750	fol. 5r (glosses)	Bronner (2013: 54–55) [= Wb]
Milan, Bibliotheca Ambrosiana, C 301 inf.	ca. 800	fol. 52v, fol. 55v marg. (narrative glosses)	Bronner (2013: 28–28) [= Ml]

"groupe rythmique", see, e.g., Wehr (2005: 356–359) and Poppe (2016: 69) for further references.

Location	Date	Passage(s) analysed	Reference
St Gall, Stiftsbibliothek, Cod. Sang. 904	ca. 850	pp. 2–3 (glosses), marginalia pp. 112, 203–204, 229 (poems)	Bronner (2013: 46–48) [= Sg]
St Gall, Stiftsbibliothek, Cod. Sang. 1395	9th century	p. 419	Bronner (2013: 48–49)
Codex Sancti Pauli, St Paul (Carinthia), Benediktinerstift, Cod. 86b/1	9th century	fol. 1v, fol. 8v	Bronner (2013: 49–50)
Book of Armagh, Dublin, Trinity College, MS 52	ca. 807	fol. 17rb–18ra	Bronner (2013: 11–13) [= Add]
Stowe Missal, Dublin, Royal Irish Academy, D ii 3 / 1238	ca. 800	fol. 65v, fol. 66r	Bronner (2013: 10–11) [= Stowe]

3.1 Word separation in Old Irish: prepositions and/or articles in combination with nouns

The present discussion focusses on nominal phrases with a (stressed) noun in combination with one or two other (unstressed) elements in three constructions: preposition plus noun (construction A), article plus noun (construction B), and preposition plus article plus noun (construction C), since these constructions both yield a sufficient number of examples in the corpus and allow comparison with the Old High German and Old Saxon data with regard to syntax.

In construction A, the largest sample in our corpus, the statistics show a clear preference for no space or minimal space between preposition and noun, i.e., *hicælech* (prep. (*h)i* + *cælech* 'into a chalice', Stowe 65v9) versus *hi linannart* (prep. *(h)i* + *linannart* 'in a linen sheet', Stowe 65v8).[6]

6 Percentages have been rounded in order to arrive at 100% overall.

Tab. 2: Spaces after prepositions

No space	Minimal space	Space	n
141	38	25	204
69%	19%	12%	

Construction A1 is a variant with prepositions triggering nasalisation, and here spaces could be inserted before and/or after the nasal indicating nasalisation, i.e., *i n hiris* (prep. *i* + nas. + *(h)iris* 'in faith', Wb 5b22)[7] or *in hétt* (prep. *i* + nas. + *(h)étt* 'into jealousy', Wb 5a13),[8] versus *iniriss* (prep. *i* + nas. + *iris* 'in faith', Wb 5b20). Modern editorial convention separates the preposition from the noun carrying the marker of nasalisation, thus privileging lexical words, e.g. *i n-iriss*. The statistics again show that no space or minimal space is the option preferred by scribes.

Tab. 3: Spaces after prepositions with nasalisation shown

	No space	Minimal space	Space	n
Before marker of nasalisation	12	3	1	16
	75%	19%	6%	
After marker of nasalisation	12	4	0	16
	75%	25%	–	

Combinations of the article with a noun, construction B, may result in *Indoblæ* (art. + *oblæ* 'the Host', Stowe 66r9) or in *In fobdod* (art. + *fobdod* 'the submersion', Stowe 66r13), the latter being the modern convention. Again, no space or minimal space emerges as the most frequent scribal choice.

7 The space in the gloss before the marker of nasalisation here appears to be conditioned by a descender from the main text. – For a study of (non-)separation of nasalisation markers in different grammatical constructions as found in the Book of Armagh see Bronner (2016).
8 In both examples there is only a minimal space after the marker of nasalisation.

Tab. 4: Spaces after article

No space	Minimal space	Space	n
62	24	14	100
62%	24%	14%	

Some case-forms of articles trigger nasalisation, construction B1, and perhaps unsurprisingly, spellings without spaces are preferred, e.g., *nanaithisse* (art. + nas. + *aithisse* 'of the insults', Stowe 66r8), versus *inna napstal* (art. + nas. + *apstal* 'of the Apostles', Stowe 66r26).

Tab. 5: Spaces after article with nasalisation shown

	No space	Minimal space	Space	n
Before marker of nasalisation	7	0	1	8
	87,5%	–	12,5%	
After marker of nasalisation	8	0	0	8
	100%	–	–	

In Old Irish, the article combines with prepositions;[9] in construction C, scribes may insert spaces after the preposition and the article, i.e., *is In charcair* (prep. + art. + *charcair* 'in the prison', Sg, p. 229, *marg. sup.*) or *forsin chombug* ([prep. + art.] + *chombug* 'on the confraction', Stowe 66r19), versus *dinchlaind* (prep. + art. + *chlaind* 'from the family', Add 17rb37), the latter being the most frequent option. Modern editors implement *din chlaind*.

Tab. 6: Spaces after preposition plus article

	No space	Minimal space	Space	n
After preposition	13	2	1	16
	81%	13%	6%	

9 For the forms see Thurneysen (1946: 293–294), there are forms of the article with or without initial *s*, depending on the preposition, e.g., *issin* 'into the' (prep. *i* 'in(to)'), but *din* 'from the' (prep. *di* 'from').

	No space	Minimal space	Space	n
After article	6	5	5	16
	38%	31%	31%	

Overall then, Old Irish scribes preferred to insert no spaces, or minimal spaces only, in the three constructions presented here. Phrases consisting of preposition plus article plus noun show the strongest deviation from this, with a space separating the unstressed combination of preposition and article from the stressed noun in about one third of the examples.

Tab. 7: Spaces after prepositions/article/preposition plus article, consolidated

No space	Minimal space	Space	n
261	56	47	364
72%	15%	13%	

These quantitative results validate scholars' conventional wisdom and their intuition that a stress group consisting of preposition plus noun, article plus noun, or preposition plus article plus noun arguably constituted a basic orthographic and conceptual unit in medieval Irish scholars' minds.

3.2 Cases and Words in *Auraicept na nÉces* 'The Scholars' Primer'

In the second part of this section, we will relate these findings to medieval Irish scholars' intuitions on nominal cases and "words" as developed in *Auraicept na nÉces* 'The Scholars' Primer'. This is the name given to a tract in which miscellaneous linguistic and grammatical topics are discussed for the professional education of poets. It may have originally consisted of a short core text, tentatively dated to a fairly early stage of the Old Irish period (ca. AD 700–900); to this core a growing body of commentary and interpretation continued to be added. The earliest manuscripts extant date to the fourteenth century. Important in our context are the *Auraicept*'s nominal, as well as its more rudimentary verbal, paradigms.[10]

Auraicept na nÉces arrives for Irish at a far larger number of nominal cases than the six cases traditionally established for Latin. This is achieved by combining the noun with various unstressed elements, mainly prepositions, but also the vocative particle, the copula, and an interrogative particle. Examples of combinations of prepositions with the masculine noun *fer* 'man' include:[11]

(1.1) aḟochsal: *óḟiur* [prep. *ó* + dat.sg.] / *óḟeraib* [*ó* + dat.pl.].
'its ablative: *a uiro* / *a uiris*.'
(1.2) ainotacht: *ifer* [prep. *i* + acc.sg.] / *ifiru* [*i* + acc.pl.].
'its illative: *in uirum* / *in uiros*.'

Thus, representatives of our construction A (preposition plus noun) are assigned specific slots in the nominal paradigm, and are (typically) written without word separation in the late medieval manuscripts.

A few examples of the combination of a noun with the unstressed article, as in our construction B, are also included in the paradigms:

(2.1) aṡelbad: *nafer* [art.gen.pl. + gen.pl.].
'its genitive: *uirorum*.'
(2.2) ainchosc: *infer* [art.acc.sg. + acc.sg.] / *innafiru* [art.acc.pl. + acc.pl.].
'its accusative: *uirum* / *uiros*.'
(2.3) athuistide: *indḟir* [art.gen.sg. + gen.sg.].
'its possessive: *uiri*.'

Finally, there is one example in the paradigms of a phrase with a preposition plus article and a noun, as in our construction C:

(3) athórmach: *frisinfer* [prep. *fri* + art.dat.sg. + dat.sg.] / *frisnaferaib* [*fri* + art.dat.pl. + dat.pl.].[12]
'its augmentative: *contra uirum* / *contra uiros*.'

---

10 For an introduction to *Auraicept na nÉces* and its transmission see Ahlqvist (1983: 11–34).
11 Presentation adapted from Ahlqvist (2000b: 611–612), the translations are Ahlqvist's. See also Ahlqvist (1983: 52–53).
12 This example appears to belong linguistically to the (late) Middle Irish period (ca. AD 900–1200). In Old Irish, the preposition *fri* is always followed by the accusative, whereas in Middle Irish it may also be followed by dative forms. While the form *fer* may be analysed both as Old Irish acc.sg. and Middle Irish acc./dat.sg., the form *feraib* is unambiguously dat.pl.

Ahlqvist (2000b: 612–613) succinctly explains how these phrases came to be assigned specific case slots and grammatical terms within the nominal paradigms, focusing on the intimate relation between grammatical categories, scribal practice, and Old Irish stress patterns:

> [... T]he view of language that underlies these paradigms is entirely consistent with Old Irish scribal practices in respect of word separation. These provide for spaces to be written, not between words as we know them nowadays, but between stress groups that correspond fairly well to the major constituents in a sentence. Thus, in the case of the 'ablative' *óḟiur*, modern practice (cf. Thurneysen 1946: 24—25) would be to write the preposition *ó* separate from the dative singular form *fiur* 'man', as follows: *ó ḟiur*. Likewise, in the 'illative' *ifer*, the preposition *i* (with the accusative it means 'into' and with the dative 'in') is nowadays usually written separate from the accusative singular *fer*, as follows, *i fer*. However, in the case of an edition of the *Auraicept* paradigms, even a modern editor must follow the early scribal practice, given the fact that it is part of an integrated framework for describing and writing the language.[13]

The scholars responsible for *Auraicept na nÉces* did not explicitly discuss their notion(s) of the grammatical concept "word", so that their ideas need to be reconstructed on the basis of their grammaticographical practice, as in the case of the nominal paradigms, or of their use of words for "word". One of these is *focal*, with the main dictionary meanings "word, phrase". This is the term used in the account of the longest word in the Irish language, which is said to have eight syllables: "Ocht sillaba isan focul as mo isan Gaedilc" (Calder [1995] 1917: 110.1435, 'Eight syllables in the longest word in Irish'). One of the two examples provided is given as

(4.1) anrocomrai[rc]nicsiumairne

in the Book of Ballymote (BB; Calder 1917: 110.1436) and as

(4.2) inrocomraircnigsiomairne

13 Modern normalisation of our constructions B and C results in *in fer* and *frisin fer* respectively. For an earlier account, which privileges syntactic units as "graphemic 'word[s]'", see Ahlqvist (1974: 185). See further Ahlqvist (1983: 29–31, 2000a: 84), Thurneysen (1928: 286–287), Hayden (2011: 10), and Tunbridge (1992: 92): "the graphical evidence of the manuscripts confirms the evidence of the *Auraicept* in showing us that the Irish preferred to think of their language in terms of syntactic clusters rather than eight (or any other number of) grammatically distinct 'parts of speech'".

in the Yellow Book of Lecan (YBL; Calder 1917: 241.4523–4524), two of the manuscripts in which the *Auraicept* is transmitted.[14] This is a verbal form, a first plural preterite of *comroircnigid* 'errs, makes mistakes', •*comraircnicsiumair*, with stress on its first syllable.[15] It is here combined with two proclitics, the augment *ro* marking past tense, and probably the prepositional relative meaning 'in which' (or perhaps the demonstrative relative meaning 'that which') and an optional enclitic first person plural "emphasising pronoun" (or *nota augens*) — in modern normalised orthography:

(4.3) an-ro·comraircnicsiumair-ne
'in which we have erred'.[16]

For an understanding of medieval Irish scholars' concepts of the word it is significant that the term *focal*, often translated as 'word', is used here to denote a stress group, consisting of a stressed verb with three unstressed clitics, parallel to *ó·ḟiur*, *i·fer*, *in·fer*, and *frisin·fer* in the paradigms of the *Auraicept* discussed above. There is another close analogue in the (less developed) verbal paradigms in the *Auraicept*. Old Irish has "passive" formations which consist of an impersonal verbal form with a clitic object pronoun, which in the case of a simple, non-compounded verb needs to be supported by a proclitic particle *no*, thus *notcartarsu* ('one loves you/you are loved') and *noncartharni* ('one loves us/we are loved'), quoted here from Ahlqvist's (2016: 103) diplomatic transcript of the relevant slot of the *Auraicept*'s verbal paradigms in Dublin, Trinity College, MS E 3. 3 (143). -*su* and -*ni* are the optional enclitic emphasising pronouns of the second person singular and first person plural respectively, supporting the

14 The other octosyllabic word is a nominal compound, more easily compatible with modern concepts of a lexical or orthographic word: *fiannamailcecheterdarai* (BB, v.ll. *fiannamailcecheterdaai* BB, *fianamailecharadhartai* YBL, Calder 1917: 110.1435–1436, 134.1740, 241.4523), which has been interpreted by Thurneysen (1928: 277) as a compound "Kriegerschaftsfreundschaften".

15 The use of the raised dot is a modern orthographic convention to indicate the position of the stressed syllable in complex verbal forms: the syllable following it carries the stress. Applied to the examples from the nominal paradigms this would result in *i·fer*, *in·fer*, and *frisin·fer*.

16 See also Calder (1917: xlvii): "These eight syllables are held to form one word. According to our present grammatical methods the basis or unity is the compound word of five syllables *comroircnigsemmar*. It is preceded by a relative pronoun *an*- and by an enclitic or pre-verb *ro*-, and it is followed by an emphasising pronominal suffix -*ni*. But the native Irish grammarians regarded all these syllables as parts of one word, and the scribes wrote the whole as one word. In their opinion proclitics were not separate words".

corresponding infixed object pronouns. In modern normalised orthography, these verbal forms are presented as *no-t·charthar-su* and *no-n·carthar-ni* respectively.

We have seen that Old Irish scribes had a tendency to write as orthographic units certain groups consisting of a stressed noun and unstressed preposition and article and that Old Irish scholars assigned to the same groups slots within the nominal paradigms as case forms in the *Auraicept*. Furthermore, the identification of a verbal form with various clitics as an example of the longest "word" in Irish is complemented by the inclusion of similar verbal forms with clitics in the verbal paradigms of the *Auraicept*. All this suggests that in the minds of Old Irish scribes and scholars the stress group was equated with an orthographic unit and, arguably, with an orthographic "word".

It needs to be emphasised, however, that Old Irish scholars were also aware of words as lexical entities, as a result of their acquaintance with the doctrines of Latin grammarians,[17] and so they applied the term *focal* in the *Auraicept* to such lexical words as well, for example in relation to *sechtae* ('heptad, sevenness'), *grus* ([type of] 'cheese'), *cloch* ('rock'), and *lind* ('pool') in the following two passages from the *Auraicept*:

(5.1) Coitchend ⁊ diles ⁊ ruidhles conadar don focul is *seachta* (Calder 1917: 56.745–746)
'Common and proper and specific are asked of the word *sechtae* ['sevenness']'

(5.2) … leithi i foclaib .i. *grus* ⁊ *cloch* ⁊ *lind*, ni fil a fregra sin lasin Laitneoir (Calder 1917: 82.1080–1082)
'… wider [is Irish] in words, i.e., *grus* [(type of) 'cheese'], *cloch* ['rock'], and *lind* ['pool'], the speaker of Latin has no equivalent for these'

The combined evidence presented here invites the interpretation that medieval Irish scribes and scholars had inevitably pre-theoretical and sometimes conflicting concepts of words as units of the grammar and the lexicon and as units of the utterance and of the prosodic hierarchy respectively, since both units can be realised as orthographic words.

17 See Parkes (1991: 3–4), quoted above, and, for example, Law (1997: 263) for some relevant concepts and terms.

4 Old High German and Old Saxon

Old High German, usually dated ca. 750–1050, provides the oldest attestations of German. In comparison with Old Irish, considerably more Old High German records exist. While some of the oldest attestations are glosses, providing not much linguistic material, the first manuscripts transmitting larger amounts of Old High German data appear at the end of the eighth century. The corpus of Old Saxon (Old Low German), the northern sibling of Old High German, is much smaller, but quite comparable generally as far as its overall characteristics are concerned.

As noted by Voetz (2006: 51), separation and non-separation of words is usually not treated in grammars and handbooks of Old High German, although some interesting phenomena have been commented on in the literature (cf. Fleischer 2009: 177–179). Importantly, to date, no empirical data have been assembled that would go beyond unsystematic observations.

The present study tried to cover as broad a scope as possible with respect to the various sources of Old High German and Old Saxon, but was limited to relatively early attestations. Chronologically, our corpus begins with the *Abrogans*, a late-eighth century codex containing an alphabetical Latin-Old High German vocabulary. Otherwise we only analysed manuscripts containing consecutive texts: prose is represented by the *Tatian*, a bilingual containing a Latin Gospel Harmony together with an Old High German translation, whereas verse texts are represented by the *Muspilli*, a partially alliterating poem that is said to display archaic features, and by Otfrid's *Liber Evangeliorum* and the *Heliand*, both vernacular renderings of the life of Jesus (alliterating in the case of the Old Saxon *Heliand*, using end rhyme in the case of the Old High German Otfrid).

Table 8 gives a survey of the manuscripts and texts analysed for the Old High German and Old Saxon data:

Tab. 8: Old High German and Old Saxon sources

Text and location	Date	Passages analysed
Abrogans St Gall, Stiftsbibliothek, Cod. Sang. 911	end of 8th century	pp. 4–6, 8–9 [partially]
Tatian St Gall, Stiftsbibliothek, Cod. Sang. 56	second quarter of 9th century	pp. 25, 157–158, 207, 219–220, 270–271

Text and location	Date	Passages analysed
Heliand M Munich, BSB, Cgm 25	ca. 850	fol. 8v, 15r, 16v (verses 522–555, 983–1009)
Heliand PL Berlin, DHM, R 56/2537	ca. 850	(verses 984–1006)
Heliand S Munich, BSB, Cgm 8840	ca. 850	fol. 3r (verses 534–558)
Otfrid V Vienna, ÖNB, Cod. 2687	last third of 9th century	fol. 71r–v, 176v–177r (verses II 23,27–24,27, V 20,30–20,71)
Otfrid P Heidelberg, UB, Cpl 52	last third of 9th century	fol. 73v–74r, 182v–183r (verses II 23,30–24,37, V 20,33–20,74)
Muspilli Munich, BSB, Clm 14098, fol. 61r, 119v–121v	end of 9th century	fol. 61r, 119v–121v (= entire record)
Otfrid D Wolfenbüttel, HAB, Cod. 131.1 Extr.	ca. 975	fol. 69v–70r (verses II 23,24–24,34)
Heliand C London, BL, MS Cotton Calig. A. VII	2nd half of 10th century	fol. 24v, 36r (verses 535–554, 985–1001)

The fact that some Old High German and Old Saxon texts are extant in different manuscripts can be exploited when analysing cases of (non-)separation. This allows first of all the comparison of the same passage in different manuscripts to see whether there is some consistency in usage as far as (non-)separation is concerned. Furthermore, we even happen to have the same text written twice by the same scribe, as in the case of the Viennese and Heidelberg Otfrid manuscripts. In the case of the *Abrogans* and the *Tatian*, the same scribe wrote the same linguistic material in Latin and Old High German, which allows comparisons between Latin and vernacular usage. In these cases, the records were used as parallel corpora: instructive insights can be gained from a comparison of different attestations of the same text.

The most important finding for the Old High German and Old Saxon records is that word separation is already well established, notwithstanding some deviations from this convention. Clearly, *scriptio continua* is not the rule in Old High German and Old Saxon writing. This can be deduced from Table 9, which gives the overall number of (non-)separation between words in relation to our threefold distinction:

Tab. 9: Spaces in Old High German and Old Saxon

No space	Minimal space	Space	n
677	1138	2418	4233
16%	27%	57%	

As Table 9 illustrates, in 57% of all cases there is a clear space; a space is lacking in only 16% of all instances, i.e., roughly in one sixth of the data. As a matter of fact, however, different parts of speech display quite different profiles with respect to their graphic expression.

4.1 Prepositions and articles/demonstrative pronouns in combination with nouns

Among the 410 Old High German and Old Saxon prepositions[18] in the corpus for which it was possible to measure the distance from the following word, no space was observed in many cases, as illustrated in Table 10:

Tab. 10: Spaces after prepositions

No space	Minimal space	Space	n
171	123	117	410
42%	30%	28%	

Note that instances of no space are significantly more frequent here than in the overall survey (42% as opposed to 16%), whereas the unclear cases of minimal spaces similarly amount to 30%, as opposed to 27% overall (see above, Table 9). If the word class of the words that are written together with the preposition is taken into account it turns out that the definite article and/or demonstrative pronoun[19] is written together with a preceding preposition in 60% of all occur-

18 Note that for one lexeme, namely OHG *zi*, OS *ti* 'to', we did not distinguish its use as a preposition from other functions. The different functions, however, do not appear to be associated with a different behaviour (cf. Busch and Fleischer 2015: 574–575, note 6).

19 Note that for the stage of German that we are dealing with here it is unclear whether grammaticalisation of the demonstrative pronoun into a definite article has already been completed;

rences, i.e., in more than half of the cases, followed by nouns (45%) and possessive pronouns (42%), whereas other parts of speech are joined to a preceding preposition much less frequently (cf. Busch and Fleischer 2015: 576, Table 5). Interestingly, there are also clear differences between individual prepositions, as illustrated by Table 11 (which is restricted to prepositional lexemes that are attested at least 30 times in our material):

Tab. 11: Spaces after different prepositions (n > 30)

	No space	Minimal space	Space	n
za/ze/zi	72%	20%	8%	65
in	50%	31%	19%	120
an	44%	31%	25%	32
mit/mid	30%	34%	36%	61
fona/fon	23%	41%	36%	69

Table 11 shows that there seems to be a significant degree of lexicalisation: While the preposition *za/ze/zi* is written together with the following word in almost three quarters of all cases, for *fon(a)* this holds true for only one quarter of cases. Interestingly, word separation seems to depend on the length of the preposition: Prepositions consisting of two (or more) syllables are much less often joined to the following word than monosyllabic ones, namely, in only ca. 21% of all cases (cf. Busch and Fleischer 2015: 577, Table 7). In sum, while our empirical study confirms that prepositions in Old High German and Old Saxon are indeed more often written together with the following word than not, there are important differences depending on the lexeme.

Table 12 surveys (non-)separation after the pronoun *der/diu/daz* (as indicated in note 19, it is unclear whether this pronoun, demonstrative in origin, has already completely grammaticalised into a definite article in the stage of Old High German – and Old Saxon – relevant to us here).

therefore, we will usually speak of the 'demonstrative pronoun/article', in order to designate OHG *der/diu/daz* in attributive usage.

Tab. 12: Spaces after *der/diu/daz*

No space	Minimal space	Space	n
13%	41%	47%	187

As can be seen, non-separation, in only 13% of all instances, is clearly rarer here than with prepositions. Interestingly, the same lexeme if employed as a relative pronoun, is much more often written together with the following word: the enclitic particle *dar*/*de*, often used after the relative pronoun, is to a large extent responsible for this effect (cf. Busch and Fleischer 2015: 585, Table 17).

In summary, prepositions are much more often joined to words following them than is the pronoun *der*/*diu*/*daz* in attributive usage. With the pronoun, the number of syllables seems to be irrelevant. Most importantly, it seems that another factor, connected to "phonetic weight", is really decisive: as the prepositions illustrate, non-separation occurs most frequently when the lexeme in question consists of two letters only (cf. Busch and Fleischer 2015: 577–578), which, in the cases considered, would correspond to two segments on the phonological level. Note that almost all forms of the pronoun *der*/*diu*/*daz* consist of at least three letters (corresponding to three phonemes in most instances).

4.2 Differences between scribes

Separation and non-separation interestingly depends on scribes and, even, individual records. For the entire corpus, the number of instances of non-separation varies between 3% and 42% (cf. Busch and Fleischer 2015: 587, Table 20). Note that these differences cannot be interpreted in terms of time or region. It is neither the older attestations nor the records originating from the same scriptorium that cluster. Applying temporal or regional criteria will yield no sensible pattern. Interestingly, we can even discern differences between the works of one scribe, as a comparison between the Viennese and the Heidelberg manuscript of Otfrid's *Liber Evangeliorum* reveals. Of all the scribes involved on these two manuscripts, two scribes, Scribe 1 and 2, were involved in the production of both the Viennese and the Heidelberg manuscript. A comparison between two passages in the two manuscripts each written by one of these scribes reveals first of all that non-separation is more frequent in the Heidelberg manuscript, the trivial reason for this being the fact that the Heidelberg manuscript provides much less room in comparison with the Viennese one. Consistent scribal patterns can be identified nevertheless. Independent of the individual manuscript, scribe 1 employs non-separation less frequently (8% in the more spacious Viennese manuscript, 16% in the smaller Heidelberg manuscript) than his colleague scribe 2 (27% in the Viennese manuscript, 42% in the Heidelberg manuscript; cf. Busch and Fleischer 2015: 587–588).

As Table 13 shows, differences between scribes cannot only be observed on a general level, but even for specific parts of speech: it shows the frequency of (non-)separation after prepositions, as implemented by different scribes, in ascending order:

Tab. 13: Spaces after preposition, as implemented by different scribes

	No space	Minimal space	Space	n
Tatian, scribe ζ	9%	–	91%	11
Abrogans	13%	63%	25%	8
Heliand PL	14%	7%	79%	14
Tatian, scribe α	17%	4%	79%	24
Heliand M	23%	54%	23%	39
Tatian, scribe γ	23%	15%	62%	26
Otfrid V, scribe 1	25%	42%	33%	36
Tatian, scribe ε	38%	54%	8%	13
Otfrid V, scribe 2	39%	54%	7%	28
Heliand C	42%	35%	23%	26
Otfrid P, scribe 1	44%	47%	9%	34
Muspilli	48%	25%	28%	65
Otfrid D	63%	29%	9%	35
Otfrid P, scribe 2	72%	24%	3%	29
Heliand S	77%	18%	5%	22

While some scribes only rarely join prepositions to the following word, others do so in nearly three quarters of instances. Note again that regional or temporal factors cannot explain the distribution. Thus, observations and generalisations regarding the (non-)separation of words do not relate to “Old High German/Old Saxon” as such; rather, first and foremost they relate to individual scribes or even individual scribal performances, as seen in the performance of the same scribes in the Viennese and Heidelberg Otfrid manuscripts respectively. Although patterns can be discerned here as well, with scribe 1 displaying a lesser amount of non-separation than scribe 2 for each manuscript, it is clear that the different manuscripts display a differing behaviour as to non-separation.

4.3 Latin vs. vernacular usage

For early medieval scribes writing their vernacular must have been a rare exception, given that the overwhelming majority of early medieval texts that have come down to us is in Latin. In that respect the German data are exceptional, and it is useful to compare them to Latin data, especially in view of the differences between vernacular and Latin usage observed by Parkes (1991: 3–4) for medieval Irish scribes quoted above. However, since the differences between individual scribes are so considerable, a comparison of a random selection of German texts with a random selection of Latin texts will not be instructive. Ideally, work by the same scribe in the same manuscript should be compared. The Old High German corpus fortunately provides material suitable for such comparisons. Both the *Abrogans* and the *Tatian* contain parallel material in both Latin and Old High German written by the same scribe. Table 14, arranged according to each individual scribe, presents the results of such a comparison:

Tab. 14: Spaces in Latin and Old High German according to individual scribes

		No space	Minimal space	Space	n
Abrogans	LAT	4%	3%	93%	100
	OHG	6%	4%	90%	101
Tatian, scribe α	LAT	8%	22%	70%	204
	OHG	4%	13%	83%	238
Tatian, scribe ε	LAT	12%	37%	51%	271
	OHG	15%	30%	55%	326
Tatian, scribe γ	LAT	5%	10%	85%	202
	OHG	3%	5%	92%	240
Tatian, scribe ζ	LAT	1%	6%	93%	173
	OHG	8%	14%	78%	219

Interestingly, this comparison reveals that the individual scribes’ treatment of Latin and Old High German is quite similar, whereas the differences between them as individuals are more considerable (cf. Busch and Fleischer 2015: 592–

593).[20] This general pattern is also replicated in rather more specific cases. A preposition *in* occurs in both Latin and German, displaying similar syntax and semantics, roughly meaning 'in(to)' in both languages, with the case of the noun phrase depending on the expression of direction or locality. As a matter of fact, in both Latin and Old High German the number of instances of non-separation of this preposition is more or less identical (31% and 32%, respectively; cf. Busch and Fleischer 2015: 593, Table 24). However, a significant difference between Latin and Old High German can be observed for the writing of the negation particle, which corresponds to significant syntactic differences between the negation in Latin and Old High German (cf. Busch and Fleischer 2015: 594, Table 25).

5 Discussion

The late antique *scriptio continua* is a principle no longer followed in the early medieval vernacular attestations analysed in our project. Old Irish scribes tend to realise stress groups as graphic units. Although these usually contain more than one syntactic word, from a phonological point of view they might be viewed as one phonological word, which thus turns out to be a fundamental unit for early medieval scribes in writing down their vernacular. Medieval Irish grammarians similarly considered phonological words consisting of a preposition plus (article plus) noun to represent specific case-forms within their system of the nominal paradigms. These semantically defined cases had the same status as the traditional cases (nominative, etc.) consisting of a bare syntactic word.[21] It needs to be kept in mind, however, that medieval Irish scholars knew the works of Latin grammarians and their ideas about "words", and the concept of the word as a unit with a specific semantic content is also reflected in *Auraicept na nÉces*.

As to Old High German and Old Saxon, word separation clearly dominates in our corpus. Some exceptions are attributable to linguistic factors. Small elements are often joined to another word; in the material discussed here, this applies to entities that are cross-linguistically known to be realised often as

20 This observation is to be verified in future studies. Given the various parameters, the statistical population is too small for statistically relevant testing of additional criteria.

21 Datives in Old Irish are rarely used on their own, i.e., without a preposition. This may have provided an additional motivation for setting up semantically defined cases consisting of preposition plus noun, beyond their status as a stress group.

clitics. Specifically, prepositions consisting of only two letters, corresponding to two phonemes, are very often written together with the following word. Another exception can be attributed to unpractised scribes who are responsible for some *scriptio continua*-like passages (cf. Busch and Fleischer 2015: 594–595). In the eighth-/ninth-century German material, these point to a lack of individual skill, not to general practice. If a "development" from *scriptio continua* to word separation is to be seen, this holds in an ontogenetic, not phylogenetic sense only. The comparison between the same scribes' performances when writing Latin and Old High German reveals that they obey roughly the same principles for the two languages (with negation as an important exception), indicating that these scribes' "grammatical reflection", insofar as it influences their writing, is the same for both Latin and the vernacular.

In summary, the (non-)separation of words in early medieval manuscripts provides important and interesting data, but its interpretation is intricate. Nevertheless, clear connections between scribal practice and the spoken language can be discerned.

References

Ahlqvist, Anders. 1974. Notes on 'case' and word-boundaries. Ériu 25. 181–189.

Ahlqvist, Anders. 1983. The early Irish linguist: An edition of the canonical part of the Auraicept na nÉces with introduction, commentary and indices (Commentationes Humanarum Litterarum 73). Helsinki: Societas Scientiarum Fennica.

Ahlqvist, Anders. 2000a. The 'native Irish grammarian' revisited. In Geraint Evans, Bernard Martin & Jonathan Wooding (eds.), Origins and revivals: Proceedings of the first Australian conference of Celtic studies, 81–90. Sydney: Centre for Celtic Studies, University of Sydney.

Ahlqvist, Anders. 2000b. The Latin tradition and the Irish language. In Sylvain Auroux, E. F. K. Koerner, Hans-Josef Niederehe & Kees Versteegh (eds.), History of the language sciences: An international handbook on the evolution of the study of language from the beginnings to the present, Vol. I, 610–614. Berlin & New York: Walter de Gruyter.

Ahlqvist, Anders. 2016. The verbal paradigms in Auraicept na nÉces. In Deborah Hayden & Paul Russell (eds.), Grammatica, gramadach, and gramadeg: Vernacular grammar and grammarians in medieval Ireland and Wales, 101–111. Amsterdam & Philadelphia: John Benjamins.

Baayen, R. Harald. 2014. Analyzing linguistic data: A practical introduction to statistics using R, 7th edn. Cambridge: Cambridge University Press.

Bischoff, Bernhard. 1990. Latin palaeography: Antiquity and the middle ages. Cambridge: Cambridge University Press.

Bronner, Dagmar. 2013. Verzeichnis altirischer Quellen, vorläufige Version. Marburg. https://hal.archives-ouvertes.fr/hal-01638388 (accessed 12 December 2017).

Bronner, Dagmar. 2016. Nasalierung im Buch von Armagh: Überlegungen zu altirischen Schreibkonventionen. Zeitschrift für celtische Philologie 63. 29–48.

Bronner, Dagmar & Nathanael Busch. 2015. Written apart and written together: Placing spaces in Old Irish and Old High German. In Andreas Nievergelt, Rudolf Gamper, Marina Bernasconi Reusser, Birgit Ebersperger & Ernst Tremp (eds.), Scriptorium: Wesen, Funktion, Eigenheiten. Comité international de paléographie latine. XVIII. Kolloquium St. Gallen 11.–14. September 2013, 519–530. München: Bayerische Akademie der Wissenschaften/C.H. Beck.

Busch, Nathanael & Jürg Fleischer. 2015. Zusammenschreibung im Althochdeutschen und Altniederdeutschen bis ca. 975: Ergebnisse einer paläographischen Untersuchung. Beiträge zur Geschichte der deutschen Sprache und Literatur 137. 563–598.

Calder, George (ed.). 1917 [1995]. Auraicept na n-Éces: The scholars' primer. Edinburgh: John Grant.

Fleischer, Jürg. 2009. Paleographic clues to prosody? – Accents, word separation, and other phenomena in Old High German manuscripts. In Roland Hinterhölzl & Svetlana Petrova (eds.), Information structure and language change: new approaches to word order variation in Germanic, 161–189. Berlin & New York: De Gruyter.

Hayden, Deborah. 2011. Poetic law and the medieval Irish linguist: Contextualising the vices and virtues of verse composition in Auraicept na nÉces. Language and History 54. 1–34.

Law, Vivien. 1997. Grammar and grammarians in the early middle ages. London & New York: Longman.

Owen, Ann Parry. 2016. Gramadeg Gwysanau: A fragment of a fourteenth-century Welsh bardic grammar. In Deborah Hayden & Paul Russell (eds.), Grammatica, gramadach, and gramadeg: Vernacular grammar and grammarians in medieval Ireland and Wales, 181–200. Amsterdam & Philadelphia: John Benjamins.

Parkes, Malcolm B. 1991. Scribes, scripts and readers: Studies in the communication, presentation and dissemination of medieval texts. London & Rio Grande, Ohio: Hambledon Press.

Poppe, Erich. 2016. Caide máthair bréithre 'what is the mother of a word': Thinking about words in medieval Ireland. In Deborah Hayden & Paul Russell (eds.), Grammatica, gramadach, and gramadeg: Vernacular grammar and grammarians in medieval Ireland and Wales, 65–83. Amsterdam & Philadelphia: John Benjamins.

Saenger, Paul. 1997. Space between words: The origins of silent reading. Stanford: Stanford University Press.

Stifter, David. 2010. Early Irish. In Martin J. Ball & Nicole Müller (eds.), The Celtic languages, 2nd edn, 55–116. London & New York: Routledge.

Thurneysen, Rudolf. 1928. Auraicept na n-Éces. Zeitschrift für celtische Philologie 17. 277–303.

Thurneysen, Rudolf. 1946. A grammar of Old Irish. Dublin: Dublin Institute for Advanced Studies.

Tunbridge, G. L. 1992. A study of scribal practices in early Irish and Anglo-Saxon manuscripts. Oxford: University of Oxford PhD dissertation.

Voetz, Lothar. 2006. Einige Beobachtungen zur Getrennt- und Zusammenschreibung im Althochdeutschen. In Ursula Götz & Stefanie Stricker (eds.), Neue Perspektiven der Sprachgeschichte: Internationales Kolloquium des Zentrums für Mittelalterstudien der Otto-Friedrich-Universität Bamberg 11. und 12. Februar 2005, 51–64. Heidelberg: Winter.

Wehr, Barbara. 2005. Focusing strategies in Old French and Old Irish. In Janne Skaffari, Matti Peikola, Ruth Carroll, Risto Hiltunen & Brita Wårvik (eds.), Opening Windows on Texts and Discourses of the Past, 353–379. Amsterdam & Philadelphia: John Benjamins.

Javier Caro Reina

Word-profiling strategies in Central Catalan, Itunyoso Trique, and Turkish

Abstract: This article investigates the patterns of syllable structure and phonotactic restrictions in three genetically unrelated languages in an attempt to identify the interaction of the strategies employed cross-linguistically to profile phonological words. The languages selected are Central Catalan, Itunyoso Trique, and Turkish. In these languages, the phonological word is the central prosodic category since syllable structure and phonotactic restrictions are sensitive to stress and within-word position. It will be shown that according to the patterns of these diagnostic criteria we can distinguish between stress-sensitive, distribution-sensitive, and harmonic words. The features of these different word types will be discussed within the framework of the typology of syllable and word languages.

Keywords: phonological processes, phonotactic restrictions, syllable structure, typology of syllable and word languages

1 Introduction

Recent work has been dedicated to the phonological (or prosodic) word (among others: Hall and Kleinheinz 1999; Dixon and Aikhenvald 2002). The phonological word grasps the interaction between phonology and morphology (Booij 1985; Nespor and Vogel 2007). For example, Nespor and Vogel (2007: 109) defined the phonological word in the following way:

> The phonological word is the lowest constituent of the prosodic hierarchy which is constructed on the basis of mapping rules that make substantial use of nonphonological notions. In particular, the phonological word (ω) represents the interaction between the phonological and the morphological components of the grammar.

Diverse diagnostic criteria have been put forward for detecting the domain of the phonological word (among others: Hall 1999: 3–8; Revithiadou 2011: 1216–

Javier Caro Reina: Romanisches Seminar, Universität zu Köln, Albertus-Magnus-Platz, D-50923 Köln, javier.caroreina@uni-koeln.de

DOI 10.1515/9783110542899-004

1221). These include phonological rules, phonotactic restrictions, minimality constraints, syllabification, and stress assignment rules (see Bergmann 2018, this volume for the phonological word as the domain of phonological rules and syllabification in German). Although there are detailed accounts of the phonological word in languages such as German (Wiese 1996: 65–74) and Swedish (Riad 2014: 117–132), little is known of the interaction of the diagnostic criteria from a cross-linguistic perspective.

This article is the first to approach the patterns of syllable structure and phonotactic restrictions in three genetically unrelated languages that are characterized by the centrality of the phonological word. The languages selected are Central Catalan, Itunyoso Trique, and Turkish. These languages differ with regard to the strategies employed for profiling the phonological word. Such strategies include word boundary signals and stress-related asymmetries that lend prominence to stressed syllables. The analysis will be conducted within the framework of the typology of syllable and word languages. The results obtained from the analysis will allow to classify word languages – that is, languages where the phonological word constitutes the central prosodic domain – according to the strategies favoured, thereby distinguishing between stress-sensitive, distribution-sensitive, and harmonic word languages.

The article is structured as follows: Section 2 introduces the theoretical framework of the typology of syllable and word languages. Section 3 presents the diagnostic criteria for assessing the relevance of the phonological word. Section 4 gives an account of the patters of these criteria in Central Catalan, Itunyoso Trique, and Turkish. Section 5 summarizes the results and discusses the implications derived from the study.

2 Typology of syllable and word languages

The relevance of the phonological word and the implications derived from the centrality of this prosodic category for the phonological make-up of a language have been specifically addressed by the typology of syllable and word languages, which has proved to be particularly fruitful in language typology, historical linguistics, and language variation.

The typology of syllable and word languages allows to classify languages according to the relevance of the prosodic categories of the syllable and the phonological word. On the basis of a geographically and genetically diversified sample of 34 languages, Auer (1993) observed that cross-linguistically the syllable and the phonological word constitute the central prosodic categories. In syllable languages, the syllable is the central prosodic category. The optimal syllable structure (CV) is shaped by Vennemann's (1988) Preference Laws, which include the Head Law, the Nucleus Law, and the Coda Law. In addition, the Contact Law and the Hiatus Law help to repair ill-formed syllable contacts (Vennemann 1988: 40–41, 50–55; Restle and Vennemann 2001: 1318–1319). By contrast, the phonological word is the central prosodic category in word languages. A number of diagnostic criteria have been proposed to evaluate the relevance of the syllable and the phonological word. These include syllable structure, distribution of the vowel and consonant inventory, and phonetic and phonological processes (Caro Reina and Szczepaniak 2014: 16–20). Thus, the patterns of these criteria are associated with the centrality of the prosodic domains of the syllable or the phonological word.

A word of caution, however, is that other prosodic categories may also shape the phonological make-up of a language. In this respect, Auer (1994: 61) points out that the mora, the phonological foot, and the phonological phrase are relevant in Japanese, Yidiɲ, and West Greenlandic, respectively. Nevertheless, the syllable seems to be the central prosodic category in these languages. Further examples of languages where the moraic foot is relevant are Late Old High German (Szczepaniak 2007: 150–154) and Welsh (Hannahs 2013). French is a prime example of a language where the phonological phrase grasps the interface between phonology and syntax, as postulated by Nespor and Vogel (2007: xx).

Importantly, the centrality of a specific prosodic domain is not static since it may change historically (see Szczepaniak 2007 for Old High German and Kümmel 2014 for Indo-Iranian). The distinct stages of the typological development may be retained in the dialects of a dialect group. This is the case in Alemannic and Catalan (Caro Reina 2013). In this respect, the typology of syllable and word languages has succeeded in explaining language variation and change in terms of a rise in the incidence of word-related features.

In his cross-linguistic study, Auer (1993) examined the patterns of selected features. These include word-related vs. syllable-related processes/phonotactics, processes deteriorating vs. optimizing the syllable structure, vowel reduction processes in unstressed syllables, word accent, tone, and syllable complexity. According to the scope and degree of word-related features, the author classifies the languages in the sample as unambiguous syllable languages (Fijian), non-prototypical syllable languages (Quechua), languages with no clear typological affiliation (Turkish), non-prototypical word languages (!Xóõ), and prototypical word languages (English), as illustrated in Table 1.

Tab. 1: Syllable-related and word-related features (adapted from Auer 1993: 94)

language	processes/ phonotactics	syllable structure rules	vowel reduction	word accent	tone	syllable complexity
Fijian	S	S	–	–	–	L
Quechua	S	S	(+)	+	–	L
Turkish	W	S/W	–	(+)	–	M
!Xóõ	W		–		W	M
English	W	W	+	+	–	H

H = high syllable complexity (CC…CC or more), L = low syllable complexity (C(G)…C or less), M = medium syllable complexity (in between including CG…CC), S = syllable-related, W = word-related

The table shows the typological affiliation of the languages according to the number of regularities referring to the syllable (top) or phonological word (bottom). However, the occurrence and interaction of the diagnostic criteria may vary cross-linguistically, raising the question of whether the strategies employed in languages with no clear classification such as Turkish and non-prototypical word languages such as !Xóõ may conflict with the strategies employed in prototypical word languages such as English. In other words, they may be mutually exclusive. This would force us to review our notion of word languages since the shape of phonological words may strongly depend on the strategies favoured. In syllable languages, phonological processes lead to an optimization of the syllable resulting in the universal syllable type CV. By contrast, in word languages phonological processes do not optimize the phonological word in the same way. As a consequence, we have to distinguish different word language types. In this vein, Hyman (2008: 335–336) speaks of demarcative, culminative, harmonic, metrical, minimal, phonotactic, and morpho-phonotactic words. This issue, which was addressed by Auer (2014: 3–4) and Caro Reina and Szczepaniak (2014: 19–20), will be discussed in more detail in the ensuing sections.

Crucially, the typology of syllable and word languages must be set apart from the classification of languages into syllable-timed and stress-timed (see Nespor, Shukla, and Mehler 2011 for a comprehensive overview of the rhythm class hypothesis). Stress-timed languages imply word languages. However, the opposite does not necessarily hold since syllable-timed languages may imply both syllable and word languages. This is the case when unstressed vowel reduction processes such as centralization are not accompanied by a reduction of vowel duration as in Central Catalan (Gavaldà-Ferré 2007) and Singapore English (Low, Grabe, and Nolan 2000). That is, the acoustic correlates of rhythm do not always succeed in grasping the word-centered affiliation of a language (see Caro Reina 2016: 20–26 for discussion).

3 Diagnostic criteria for the phonological word

The criteria selected for evaluating the centrality of the phonological word in Central Catalan, Itunyoso Trique, and Turkish are syllable structure (Section 3.1) and phonotactic restrictions (Section 3.2). The patterns of these criteria may give hints about the relevance of the phonological word. This is the case when they are sensitive to stress (stressed vs. unstressed syllables) and within-word position (word-initial, word-medial, and word-final). Additional criteria have been proposed in the literature. These include word-optimizing processes and demarcative stress (see Caro Reina and Szczepaniak 2014: 16–20 for details). Demarcative stress is found in Itunyoso Trique and Turkish, both of which have word-final stress. Altogether, syllable structure complexity, phonotactic restrictions related to (or profiling) the phonological word, and demarcative stress provide cues for speech segmentation (see Boll-Avetisyan 2018, this volume for details).

3.1 Syllable structure

For a classification of syllable complexity, I will follow Maddieson (2013a), who distinguishes between simple, moderately complex, and complex syllable structures, as shown in (1) (for other classifications see Auer 1993: 41–44 and Levelt and van de Vijver 2004). Simple syllable structures include the syllable types CV and V, where V represents a short vowel, a long vowel, or a diphthong. Moderately complex syllable structures allow two consonants in the onset and/or one consonant in the coda. However, the second consonant in the cluster is restricted to a liquid or a glide as in Germ. *groß* [gʀoːs] 'big'. This combination is widely distributed in the languages of the world and hence constitutes an unmarked structure (see Parker 2012 for a cross-linguistic study). In complex syllable structures, the combination of two consonants is not restricted to sequences of consonant + liquid/glide as in Germ. *schnell* [ʃnɛl] 'fast'. Additionally, they allow three or more consonants in the onset as well as two or more consonants in the coda. Complex syllables may contain ill-formed syllable structures that violate the Sonority Sequencing Principle (see Parker 2011: 1161–1167 for a comprehensive overview). This is the case with Germ. *Spiel* [ʃpiːl] 'game', where the sibilant is more sonorous than the following stop. Thus, the segment constitutes an extrasyllabic element. Syllable complexity may arise from affixation as in Germ. *mach-st* [maxst] 'make-2SG'. In this respect, Hall (2011: 220) speaks of "morphologically complex clusters". The author further observes that in languages with complex syllables syllable complexity may be higher in inflected forms than in monomorphemic forms (see Kohler 1995: 182–183 for German).

(1) Syllable templates according to syllable complexity

simple:	(C)V
moderately complex:	(C)(C)V(C)
complex:	(C)(C)(C)V(C)(C)(C)(C)

With regard to Maddieson's (2013a) classification, Kouwenberg (2010: 365–367) critically discusses the status of prenasalized stops (single consonants vs. complex consonant clusters), the occurrence of syllable complexity according to within-word position, and the markedness of the syllable type V. Additionally, there are contentious issues such as the absolute and relative frequency of complex syllable structures. In this respect, I will assess the syllable complexity of a language regardless of the frequency of complex syllables. That is, languages with a reduced set of complex syllables will be classified as languages with a complex syllable structure. This is the case in Itunyoso Trique, as we will see in Section 4.2.

In phonological typology, syllable complexity has been traditionally addressed in terms of implications (Greenberg, Osgood, and Jenkins 1966: xxv; Greenberg 1978: 248–249). For example, moderately complex syllables imply simple syllables while complex syllables imply both simple and moderately complex syllables. That is, languages with simple syllables do not necessarily have complex or moderately complex syllables. However, these implications do not give hints about the relevance of the prosodic categories of the syllable and the phonological word. Crucially, complex and moderately complex syllables may demarcate phonological words when they are associated with word and morpheme boundaries. For example, complex syllables may occur only word-initially as in Georgian, only word-finally as in Catalan, and both word-initially and word-finally as in German. Interestingly, in languages with a simple syllable structure the syllable type V may contribute to profiling the phonological word when its occurrence is restricted to word-initial position. This is the case in Kulina, an Arawan language spoken in south-western Amazonia, where onset-less syllables delimit the left margin of the phonological word (see Dienst 2014: 27–29 for details).

From a diachronic perspective, complex syllables may be inherited from a prior language stage. For example, the Italian word-initial clusters [sp st sk] are derived from Latin as in Lat. STĒLLA> It. *stella* 'star'. Alternatively, they may result from phonological processes such as unstressed vowel deletion and consonant epenthesis. Unstressed vowel deletion gives rise to complex syllable structures at word and morpheme boundaries as in MHG *gib-et* > NHG *gib-t* 'give-3SG'. Additionally, consonant epenthesis increases syllable complexity at the margins of the phonological word as in MHG *saf* > NHG *Saft* 'juice' (see Szczepaniak 2007: 249–257, 2014: 172–173 for details). In historical grammars of German, these instances of consonant epenthesis have been traditionally explained in terms of analogy (Moser 1951: 44–84; Paul 1998: 160–161). However, analogy does not succeed in explaining why the process did not operate in the opposite direction, thereby simplifying moderately complex syllable structures. Thus, the process can be explained more adequately when described as part of a typological shift.

Auer (1993: 42) addresses the question of whether syllable complexity is found at the phonemic or phonetic level. The author illustrates this issue with vowel epenthesis in Toda, a Dravidian language spoken in southern India. In Toda, underlying complex syllables are repaired by means of vowel epenthesis. As a consequence, underlyingly Toda has a complex syllable structure while on the surface it has a simple syllable structure. The analysis of the syllable structure conducted in the case studies (Section 4) is based on the phonetic level.

3.2 Phonotactic restrictions

Phonotactic restrictions may be stress-related and position-related. Stress-related restrictions help to create an asymmetry within the phonological word. This asymmetry may be achieved by increasing the number of vowels in stressed syllables. This is the case with vowel lengthening and diphthongization. Alternatively, the asymmetry may be achieved by reducing the number of vowels in unstressed syllables. This is the case with unstressed vowel reduction processes, which include shortening, simplification of diphthongs, denasalization, unrounding, centralization, and merger of back vowels (for unstressed vowel reduction processes see Dauer 1983: 57; Auer 1993: 66; Bybee et al. 1998: 280; Barnes 2006: 20). For example, in Palauan, an Austronesian language spoken in the Republic of Palau, underlyingly short vowels, long vowels, and diphthongs undergo centralization, shortening, and diphthong simplification, respectively (see Josephs 1975: 20–21, 56–57, 59–62 for details). Schwa is associated with unstressed syllables while long vowels and diphthongs are associated with stressed syllables. Similarly, Kohler (1995: 223) observes that in German, schwa is restricted to unstressed syllables, where it occurs with a frequency of 60%. That is, in German schwa can be exclusively associated with unstressed syllables. From a synchronic perspective, it is not always straightforward whether an asymmetry resulted from processes operating in stressed syllables or, rather, in unstressed syllables. This issue will be discussed in more detail in Section 4.2. In addition to unstressed vowel reduction processes, vowel harmony can help to neutralize the set of vowels licensed within the phonological word.[1] Depending on whether a word language employs unstressed vowel reduction or vowel harmony, I will speak of "stress-sensitive words" and "harmonic words", respectively.

Position-related restrictions may lead to the emergence of word boundary signals. This is the case when the occurrence of a sound (or group of sounds) can be exclusively associated with the word-initial or word-final position of the phonological word. In this respect, Trubetzkoy (1971: 242) speaks of "positive signals". In phonological typology, the occurrence of sounds has been traditionally addressed in terms of markedness and inventory size (Maddieson 2013b, 2013c). Crucially, marked and unmarked sounds do not give hints about the relevance of the prosodic categories of the syllable and the phonological word since their distribution may be either syllable-related or word-related. Let us observe the distribution patterns of click sounds, which typically occur in languages with a large consonant inventory (see Table 2). For example, in !Xóõ, a Khoisan language spoken in Botswana and Namibia, clicks are restricted to word-initial onsets. That is, their distribution is sensitive to within-word position. By contrast, in Sandawe, a Khoisan language spoken in Tanzania, clicks may occur both in word-initial and word-medial onsets. That is, their distribution is syllable-related.[2] I will speak of "distribution-sensitive words" when a word language employs strong position-related restrictions. Similar to syllable

1 In previous work (Auer 1993: 44–45, 2001: 1397–1398), vowel harmony was linked to syllable languages. However, Kabak (2014: 120–121) points out that in Turkish vowel harmony is word-related and word-optimizing since it has a demarcative function. Thus, vowel harmony may also help to profile phonological words (see Caro Reina 2016: 54–55 for discussion).

2 More specifically, Sandawe has 25 clicks. While 20 are licensed both in word-initial and word-medial onsets, only 5 are restricted to word-initial onsets. These include the aspirated dental, nasalized dental, aspirated post-alveolar, voiced post-alveolar, and voiceless lateral clicks (see Hunziker, Hunziker, and Eaton 2008: 17–23, 74–75 for details).

structure, we have to distinguish between position-related restrictions at the phonemic and phonetic level. This issue will be addressed in Section 4.2.

Tab. 2: Position-related distribution of clicks in !Xóõ and Sandawe

	Word-initial onset	Word-medial onset	Source
!Xóõ	+	–	Traill (1985: 164–166)
Sandawe	+	+	Hunziker, Hunziker, and Eaton (2008: 23)

In contrast to word languages, syllable languages do not exhibit stress-related restrictions. For example, in Spanish the same set of vowels and consonants is licensed in stressed and unstressed syllables (Quilis 1999: 144, 381; Martínez, Fernández, and Carrera 2003: 256; Szczepaniak 2009). Additionally, in syllable languages we may find position-related restrictions. However, they are syllable-related. That is, we do not need a prosodic constituent larger than the syllable in order to adequately explain their occurrence.

4 Case studies

Let us examine the patterns of the diagnostic criteria presented in the previous section in three genetically unrelated languages: Central Catalan (Section 4.1), Itunyoso Trique (Section 4.2), and Turkish (Section 4.3). The analysis will focus on monomorphemic words. Inflected words, derived words, and compounds will not be discussed.

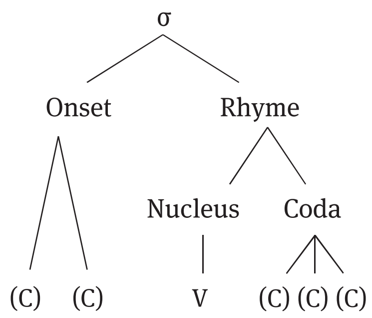
4.1 Central Catalan

Catalan is a Romance language spoken in eastern Spain (Catalonia, the Valencian Community, the Balearic Islands, and parts of Aragon and Murcia), Andorra, Roussillon, and the Sardinian city of Alghero. Catalan is traditionally divided into two main dialect groups: Western Catalan (*català occidental*) and Eastern Catalan (*català oriental*) (Veny 1991: 244–245, 2002: 19–20). Western Catalan comprises North-Western Catalan (*català nord-occidental*) and Valencian (*valencià*) while Eastern Catalan consists of Central Catalan (*català central*), Roussillon Catalan (*rossellonès*), Balearic (*balear*), and Alghero Catalan (*alguerès*).

The following account deals with Central Catalan and is based on Julià (2002), Lloret (2002), and Mascaró (2002) (see Caro Reina 2014 for details).

The syllable template of Central Catalan is presented in (2). Central Catalan has a complex syllable structure. Syllable onsets may contain up to two consonants while syllable codas may contain up to two or three consonants depending on whether they appear word-medially or word-finally, respectively. Syllable complexity mostly occurs word-finally, thereby delimiting the right margin of the phonological word. Word-final codas containing three consonants are found in monomorphemic words such as *text* [tekst] 'text' (see Lloret 2002: 219 for details and Recasens 1996: 178–179 for the combination of two and three consonants in word-final codas). However, they are more frequent in morphologically complex words. This is the case with the plural ending and the ending of the second person singular of the present indicative as in *forn-s* [fɔɾns] 'oven-PL' and *dorm-s* [dɔɾms] 'sleep-2SG.PRS.IND', respectively. Word-final clusters may violate the Sonority Sequencing Principle. Although extrasyllabic elements marginally occur in monomorphemic words such as *algeps* [əɫˈd͡ʒeps] 'gypsum', they are generally associated with morphological information as in *cap-s* [kaps] 'head-PL' and *sap-s* [saps] 'know-2SG.PRS.IND'. Complex syllables emerged from apocope and syncope, both of which applied in Pre-Old Catalan as in VLat. PǑRCO [ˈpɔɾko] > OCat. *porc* [pɔɾk] 'pig' and VLat. PǑRC-OS [ˈpɔɾkos] > OCat. *porc-s* [pɔɾks] 'pig-PL', respectively. Additionally, consonant epenthesis contributed to increasing word-final syllable complexity. The process applied both to word-final open syllables as in *api* 'celery' [ˈapi] > [ˈapit] and to word-final closed syllables as in *mar* 'sea' [maɾ] > [maɾt]. As a consequence, word-final simple syllable structures became moderately complex (CV > CVC) while word-final moderately complex syllable structures became complex (CVC > CVCC).

(2) Central Catalan surface syllable structure



Stress-related and position-related restrictions are depicted in Figure 1 (geminates and consonant clusters are not included). Note that Central Catalan has lexical stress as in *parl-ar* [pərˈla] 'speak-INF' and *parl-a* [ˈparlə] 'speak-3SG.IND'. Stress-related restrictions involve unstressed vowel reduction processes such as centralization, merger of back vowels, and simplification of diphthongs. There is a seven-to-three inventory reduction whereby underlying /a e ɛ/ undergo centralization and /ɔ o u/ merge into [u]. That is, [e ɛ a ɔ o] are associated with stressed syllables while [ə] is associated with unstressed syllables. Schwa is the most common vowel in unstressed syllables with a relative frequency of 62%.[3] Stress-related restrictions are not foot-related. This issue will be illustrated with the patterns of unstressed vowel reduction in *casolan-s* [kəzuˈlans] 'home-made-PL', which consists of the weak foot $[\text{kəzu}]_{\text{Fw}}$ and the strong foot $[\text{lans}]_{\text{Fs}}$ (see Lloret and Jiménez 2008: 63 for a representation of the phonological structure of the item). Note that centralization and merger of back vowels apply in unstressed syllables within the phonological word. Crucially, centralization applies to the prominent syllable of the weak foot. This implies that unstressed vowel reduction is not foot-sensitive. A foot-based unstressed vowel reduction process would have resulted in $*[[\text{kazu}]_{\text{Fw}}[\text{ˈlans}]_{\text{Fs}}]_{\omega}$, where only non-prominent syllables of strong and weak feet undergo reduction. Moreover, Wheeler (2005: 277) observes that secondary stress is not relevant for the prosodic organization of phonological words. The weak foot [kəzu] may thus be either trochaic ([ˌkəzu]) or iambic ([kəˌzu]). With regard to position-related restrictions, Central Catalan exhibits the positive signals [β ð ɣ v d͡z d͡ʒ], which are restricted to word-medial onsets (highlighted in bold). However, they disappear in connected speech. For example, [β ð ɣ] may also occur word-initially as a result of lenition while [v d͡z d͡ʒ] may also occur word-finally as a result of voicing across words. Hence, in Central Catalan phonotactic restrictions are exclusively associated with stress.

3 This value is based on De Yzaguirre's (1995: 91) account of the frequency of underlying vowels in unstressed syllables. More specifically, schwa has a frequency of 1,115,604 tokens resulting from the sum of the occurrence of unstressed /a/ (497,720 tokens) and /e/ (617,884 tokens).

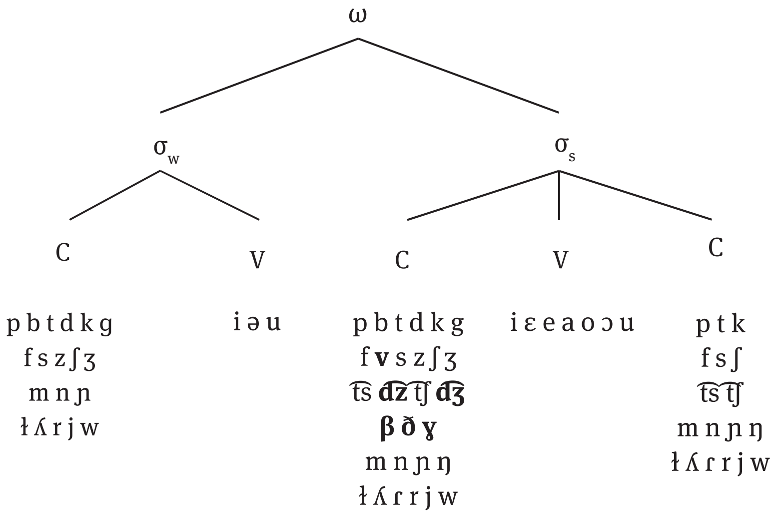


Fig. 1: Structure of disyllabic phonological words in Central Catalan

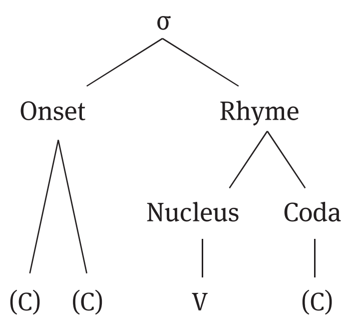
In sum, Central Catalan profiles the phonological word by means of word-final syllable complexity and stress-related restrictions. The presence of syllable complexity in stressed syllables and stress-dependent vowel reduction allows to classify Central Catalan as a stress-sensitive word language. Interestingly, Central Catalan exhibits resyllabification, which is viewed as a typical feature of syllable languages since it optimizes the syllable structure within the phonological phrase (Auer 2001: 1397; Szczepaniak 2007: 36, 52). However, the process is also found in word languages such as English and Russian, both of which were classified as unambiguous word languages (Auer 1993: 50, 94). Word-centered languages have developed different strategies for coping with resyllabification. These include glottal stop insertion as in German (Kohler 1994; Wiese 1996: 58–60) and obstruent voicing as in Catalan (Bonet and Lloret 1998: 118–122), Dutch (Booij 1999: 146–147), and Luxembourgish (Gilles 2014: 295–300) (see Lloret and Jiménez 2009 for voicing across word boundaries). This implies that resyllabification alone is not indicative of the syllable-centered affiliation of a language. Notwithstanding the presence of resyllabification, obstruent voicing helps to demarcate word and morpheme boundaries in Central Catalan (see Caro Reina 2014: 373–374 for discussion).

4.2 Itunyoso Trique

Itunyoso Trique is an Oto-Manguean language spoken in the Mexican town of San Martín Itunyoso. The sources for the analysis are based on DiCanio (2008, 2010).[4]

The syllable template of Itunyoso Trique is given in (3). Itunyoso Trique has a complex syllable structure. Syllable onsets may contain up to two consonants. Consonant clusters mostly occur word-initially. With the exception of /sw/, they violate the Sonority Sequencing Principle. This is the case with /st sk sn ʃk ʃkw rt rk rkw rm/. The preference for word-initial ill-formed syllable structures becomes evident from Spanish loan words containing sequences of stop + liquid, which are adapted as sequences of liquid + stop as in Span. *cruz* 'cross' [kɾus] > [rku^{4}si^{43}]. Word-initial consonant clusters arose from unstressed vowel deletion, which can be still observed in the alternation between [si^{3}keʔ^{3}] ~ [skeʔ^{3}] 'mud'. In native words we find open syllables. Closed syllables only occur word-finally. The consonants permitted in the word-final coda are the laryngeal segments [ʔ h] as in [ja^{3}koʔ^{3}] 'forest' and [ɾa^{3}kah^{3}] 'iguana', respectively. Thus, complex syllables are associated with the left margin of the phonological word while moderately complex syllables are associated with the right margin of the phonological word. That is, syllable complexity helps to delimit the margins of the phonological word.

(3) Itunyoso Trique surface syllable structure



In Itunyoso Trique we find stress-related and position-related restrictions, as depicted in Figure 2 (geminates and consonant clusters are not included). Note

4 I would like to thank Christian T. DiCanio for comments on Itunyoso Trique.

that Itunyoso Trique has word-final stress, which helps to demarcate the right margin of the phonological word. With regard to stress-related restrictions, short vowels (oral and nasal) occur both in stressed and unstressed syllables while long vowels (oral and nasal) only occur in stressed syllables.[5] Interestingly, there are no vowels that can be exclusively associated with unstressed syllables. In this respect, Itunyoso Trique differs from Central Catalan, where schwa is coupled with unstressed syllables. Underlyingly nasal vowels are restricted to stressed syllables. In other words, oral and nasal vowels only contrast in stressed syllables. However, nasal vowels may also surface in pretonic syllables as a result of nasal spreading as in /ju^{3}ʔũh^3/ [jũ^{3}ʔũh^3] 'woman' (see DiCanio 2008: 47–48 for details). The process is word-related since it applies within the phonological word. In word-final open syllables, vowels undergo lengthening. This implies that vowel length is not phonemic. Additionally, Itunyoso Trique is a tonal language with level, falling, and rising tones. Altogether, there are nine tones, all of which are contrastive in word-final syllables. However, there are only three and two contrasts in penultimate and antepenultimate syllables, respectively. Thus, stress-related asymmetries include vowel lengthening in stressed syllables and a reduced set of tones in unstressed syllables.

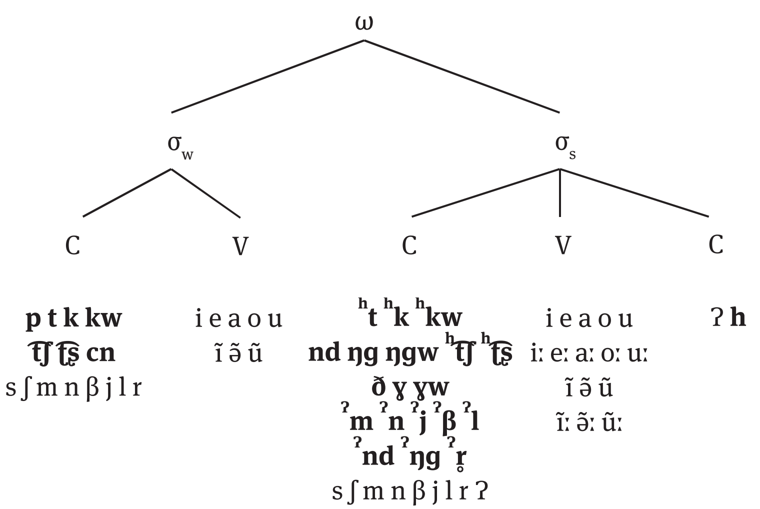


Fig. 2: Structure of disyllabic phonological words in Itunyoso Trique

5 DiCanio (2010) does not indicate vowel length in the phonetic transcriptions. However, the author supports evidence for vowel length (p. 236). For this reason, long vowels are given in Figure 2.

In disyllabic phonological words, the distribution of consonants is extremely sensitive to within-word position. For example, [p t k kʷ t͡ʃ ʈ͡ʂ cn] are restricted to word-initial onsets while [ʰt ʰk ʰkw nd ŋg ŋgw ʰt͡ʃ ʰʈ͡ʂ ð ɣ ɣw ˀm ˀn ˀj ˀβ ˀl ˀnd ˀŋg ˀr̥] are restricted to word-medial onsets. Additionally, [h] is only licensed in word-final codas. That is, these consonants constitute positive signals (highlighted in bold).[6] In contrast, the consonants [s ʃ m n β j l r ʔ] do not help to demarcate the phonological word. For example, [s ʃ m n β j l r] may occur both in word-initial and word-medial onsets while [ʔ] may occur both in word-medial onsets and word-final codas. Since word-medial positive signals include preaspirated and glottalized consonants, the question arises as to whether these asymmetries were achieved by means of deaspiration and deglottalization in word-initial unstressed syllables or, rather, by means of preaspiration and glottalization in word-medial stressed syllables. This question cannot be answered without historical or comparative evidence.

In sum, Itunyoso Trique profiles the phonological word by means of syllable complexity, stress-related and position-related restrictions, and word-final stress. In view of the strong position-related restrictions, we can classify Itunyoso Trique as a distribution-sensitive word language. Interestingly, Itunyoso Trique is a tonal language that deviates from the typological patterns expected. With regard to tonal languages, Auer (1993: 88) found the following correlations: a) tone negatively correlates with complex syllable structures, b) tone mainly occurs in languages with no or vague word stress, and c) word-related phonotactics or processes do not occur in languages with non-restricted tone. Similarly, according to Maddieson (2013d) there is a tendency for tonal languages to have less complex syllable structures. These correlations, however, are not borne out. Notwithstanding the presence of non-restricted tone, Itunyoso Trique has word-final stress, word-related phonotactics, and word-related processes such as nasal spreading.

6 Note that word-initial onsets of disyllabic words and word-medial onsets of trisyllabic words do not differ with respect to the set of consonants licensed. However, the number of consonants permitted in word-initial onsets is higher in disyllabic than in trisyllabic words (15 vs. 10). More specifically, [t k kw t͡ʃ s ʃ m n l r] may occur in word-initial onsets of disyllabic and trisyllabic words while [p ʈ͡ʂ β j cn] are absent from word-initial onsets of trisyllabic words (see DiCanio 2008: 43, 2010: 233–234 for details). This implies that there are no positive signals in the word-initial onset of trisyllabic words. A possible explanation is the low frequency of trisyllabic words. According to DiCanio (2008: 20), disyllabic and trisyllabic words have a relative frequency of 66% and 12%, respectively. In other words, positive signals are more effective in disyllabic words.

4.3 Turkish

Turkish is a Turkic language spoken predominantly in Turkey. The sources for the analysis include Lewis (1967), Clements and Sezer (1982), Zimmer and Orgun (1992), and Kabak (2011).

The syllable template of Turkish is presented in (4). Turkish has a complex syllable structure. Syllable onsets may contain one consonant while codas may contain one or two consonants depending on whether they occur word-medially or word-finally, respectively. According to Clements and Sezer (1982: 245), word-final consonant clusters involve sequences of sonorant + obstruent, fricative + stop, and stop + fricative as in *aşk* 'love', *kalp* 'heart', and *raks* 'dance', respectively. In contrast to the sequences of sonorant + obstruent and fricative + stop, the sequence of stop + fricative (*ks*) violates the Sonority Sequencing Principle.

(4) Turkish surface syllable structure

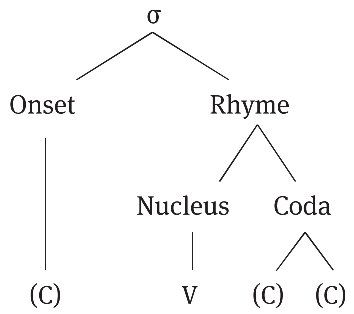


Figure 3 shows the distribution of vowels and consonants in disyllabic phonological words. Note that Turkish has word-final stress. The distribution of vowels is position-related and is constrained by vowel harmony. In Turkish we find palatal and labial harmony. In the following, I will briefly summarize both vowel harmony processes (see Clements and Sezer 1982 and Kabak 2011 for a detailed account). Turkish has eight vowel phonemes. These include /i y e ø ɑ o ɯ u/, which are licensed in word-initial syllables. However, the front vowels /i y e ø/ cannot be followed by the back vowels /ɑ o ɯ u/ and vice versa. Additionally, non-initial high vowels must agree in roundness with the word-initial vowel. That is, the vowels occurring in word-initial and non-word-initial syllables must harmonize with respect to the features [±back] and [±round] (see Kabak 2011: 2833 for examples). In this respect, front-back and rounding oppositions are neutralized in non-initial syllables. In disyllabic words, non-initial syllables are stressed while in trisyllabic words they contain a sequence of unstressed and stressed syllables. In other words, vowel harmony contributes to creating position-related asymmetries within the phonological word. Kabak (2014: 120–121) argues that vowel harmony, together with word-final stress, helps to delimit word boundaries in online processing. The distribution of consonants is not sensitive to within-word position since the set of consonants [p t k t͡ʃ s z ʃ h j m n l ɾ] may occur in word-initial onsets, word-medial onsets, and word-final codas. Thus, in Turkish we find no stress-related restrictions.

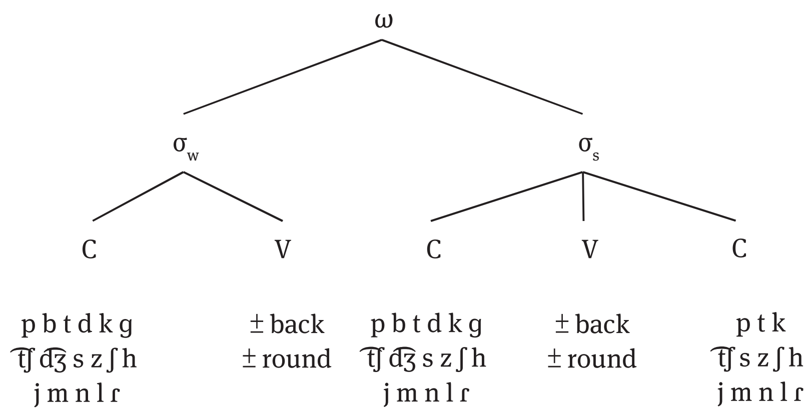


Fig. 3: Structure of disyllabic phonological words in Turkish

In sum, Turkish profiles the phonological word by means of word-final syllable complexity, position-related restrictions, and word-final stress. The presence of vowel harmony allows to classify Turkish as a harmonic word language. Interestingly, in Turkish vowel asymmetries are position-related and occur in non-initial syllables. This is due to the fact that in Turkish vowel harmony constitutes a rightward agreement operation. By contrast, in Central Catalan and Itunyoso Trique vowel asymmetries are stress-related. Additionally, in Turkish position-related restrictions involve the distribution of vowels while in Itunyoso Trique they involve the distribution of consonants.

5 Summary and discussion

This article has provided an analysis of the strategies for profiling the phonological word in Central Catalan, Itunyoso Trique, and Turkish. In view of the relevance of the phonological word gleaned from the patterns of syllable structure and phonotactic restrictions, these languages can be classified as word languages. According to the strategies employed, we can distinguish at least three types of word languages: stress-sensitive, distribution-sensitive, and harmonic. Central Catalan is a stress-sensitive word language that profiles phonological words by means of syllable complexity in stressed syllables and stress-dependent vowel reduction. Further examples of stress-sensitive word languages are Friulian, English, German, Portuguese, and Palauan. Stress-sensitive word languages are characterized by asymmetries within the phonological word. These asymmetries result from unstressed vowel reduction processes such as centralization. Alternatively, they arise from processes that enlarge the set of vowels in stressed syllables such as lengthening and diphthongization. Interestingly, in Central Catalan the imbalance between stressed and unstressed syllables is only achieved by means of unstressed vowel reduction processes. Contrary to other stress-sensitive word languages such as Friulian, vowel lengthening and diphthongization are absent from Central Catalan. Stress-sensitive word languages typically have complex syllables. This is in line with the observation that complex syllable structures positively correlate with word-related processes such as stress-dependent vowel reduction (Auer 1993: 88). Interestingly, stress-sensitive word languages do not make use of strong position-related restrictions, which are fully exploited in distribution-sensitive languages. In this respect, stress-related restrictions seem to negatively correlate with position-related restrictions.

Itunyoso Trique is a distribution-sensitive word language where the distribution of consonants is extremely sensitive to within-word position. Similar to Itunyoso Trique, !Xóõ and Khoekhoe exhibit strong phonotactic restrictions (Auer 1993: 74). For example, in word-initial onsets !Xóõ licences up to 116 phonemic contrasts including clicks, stops, fricatives, and nasals while it only licenses the six consonants /b dy m n ɲ l/ in word-medial onsets (see Traill 1985: 164–166 for details and Auer 1994: 62–64 for discussion). Notwithstanding the occurrence of position-related restrictions, we may also find stress-related restrictions. However, these do not seem to be as strong as position-related restrictions. This would explain why !Xóõ was not classified as a prototypical word language as opposed to English (see Table 1).

Finally, Turkish is a harmonic word language where the distribution of vowels is constrained by vowel harmony. In contrast to Central Catalan, where the imbalance between stressed and unstressed syllables is achieved by means of unstressed vowel reduction processes, in Turkish this imbalance is achieved by means of vowel harmony. The absence of unstressed vowel reduction processes motivated the classification of Turkish into a language with no clear typological affiliation (see Table 1). However, as Auer (1993: 88) points out, languages with stress-dependent vowel reduction lack vowel harmony. Similar to Turkish, Finnish exhibits harmonic words. In harmonic word languages, position-related restrictions involve the distribution of vowels while in distribution-sensitive languages they involve the distribution of consonants.

In addition to stress-sensitive word languages, distribution-sensitive and harmonic word languages may have complex syllables. For example, in Itunyoso Trique syllable complexity occurs word-initially while in Turkish it occurs word-finally. Additionally, both languages have clusters that violate the Sonority Sequencing Principle. Future studies are required to determine whether syllable complexity is found more frequently in stress-sensitive word languages than in distribution-sensitive and harmonic word languages. Syllable complexity may decrease at word boundaries as a result of inflection and derivation. This is the case in Central Catalan and Turkish, where the word-final complex syllables in *llarg* [ʎaɾk] 'long[M]' and *Türk* [tyɾk] 'Turk' become moderately complex in inflected forms, giving rise to *llarg-a* [ˈʎaɾ.ɣə] 'long-F' and *Türk-e* [tyɾ.ˈke] 'Turk-DAT', respectively. The features of stress-sensitive, distribution-sensitive, and harmonic word languages are summarized in Table 3.

Tab. 3: Word language types

stress-sensitive	▪ Central Catalan, English, Friulian, German, Palauan, Portuguese ▪ stress-related restrictions resulting from unstressed vowel reduction (e.g. centralization) and stressed vowel enlargement (e.g. lengthening) ▪ complex syllables
distribution-sensitive	▪ Itunyoso Trique, Khoekhoe, !Xóõ ▪ strong position-related restrictions ▪ stress-related restrictions possible ▪ complex syllables possible
harmonic	▪ Turkish, Finnish ▪ stress-related restrictions resulting from vowel harmony ▪ complex syllables possible

Abbreviations

CCat.	Central Catalan
Germ.	German
It.	Italian
Lat.	Latin
MHG	Middle High German
NHG	New High German
OCat.	Old Catalan
Span.	Spanish
VLat.	Vulgar Latin

References

Auer, Peter. 1993. Is a rhythm-based typology possible? A study of the role of prosody in phonological typology (KontRI Working Paper 21). Konstanz: Universität Konstanz.

Auer, Peter. 1994. Einige Argumente gegen die Silbe als universale prosodische Hauptkategorie. In Karl Heinz Ramers, Heinz Vater & Henning Wode (eds.), Universale phonologische Strukturen und Prozesse (Linguistische Arbeiten 310), 55–78. Berlin & New York: De Gruyter.

Auer, Peter. 2001. Silben- und akzentzählende Sprachen. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), Language typology and language universals, vol. 2 (Handbücher zur Sprach- und Kommunikationswissenschaft 20), 1391–1399. Berlin & New York: Walter de Gruyter.

Auer, Peter. 2014. Preface. In Javier Caro Reina & Renata Szczepaniak (eds.), Syllable and word languages (linguae & litterae 40), 1–7. Berlin & Boston: De Gruyter.

Barnes, Jonathan. 2006. Strength and weakness at the interface (Phonology and Phonetics 10). Berlin: Mouton de Gruyter.

Bergmann, Pia. 2018. The phonological word in German – Insights from an acoustic-phonetic study of complex words. In Christiane Ulbrich, Alexander Werth & Richard Wiese (eds.), Empirical approaches to the phonological structure of words (Linguistische Arbeiten 567), 13–44. Berlin & Boston: De Gruyter.

Boll-Avetisyan, Natalie. 2018. The role of phonological structure in speech segmentation by infants and adults: a review and methodological considerations. In Christiane Ulbrich, Alexander Werth & Richard Wiese (eds.), Empirical approaches to the phonological structure of words (Linguistische Arbeiten 567), 213–240. Berlin & Boston: De Gruyter.

Bonet, Eulàlia & Maria-Rosa Lloret. 1998. Fonologia catalana. Barcelona: Ariel.

Booij, Geert. 1985. The interaction of phonology and morphology in prosodic phonology. In Edmund Gussmann (ed.), Phono-morphology. Studies in the interaction of phonology and morphology, 23–34. Lublin: Katolicki Uniwersytet Lubelski.

Booij, Geert. 1999. The phonology of Dutch. Oxford: Oxford University Press.

Bybee, Joan, Paromita Chakraborti, Dagmar Jung & Joanne Scheibman. 1998. Prosody and segmental effect: Some paths of evolution for word stress. Studies in Language 22(2). 267–314.

Caro Reina, Javier. 2013. Phonological variation in Catalan and Alemannic from a typological perspective. In Peter Auer, Javier Caro Reina & Göz Kaufmann (eds.), Language Variation – European Perspectives IV. Selected papers from the Sixth International Conference on Language Variation in Europe (ICLaVE 6), Freiburg, June 2011 (Studies in Language Variation 14), 27–44. Amsterdam: John Benjamins.
Caro Reina, Javier. 2014. Central Catalan in the framework of the typology of syllable and word languages. In Javier Caro Reina & Renata Szczepaniak (eds.), Syllable and word languages (linguae & litterae 40), 349–390. Berlin & Boston: De Gruyter.
Caro Reina, Javier. 2016. Central Catalan and Swabian in the framework of the typology of syllable and word languages. A synchronic and diachronic account of the prosodic domains of the syllable and the phonological word. Freiburg: University of Freiburg dissertation.
Caro Reina, Javier & Renata Szczepaniak. 2014. Introduction: Syllable and word languages. In Javier Caro Reina & Renata Szczepaniak (eds.), Syllable and word languages (linguae & litterae 40), 8–40. Berlin & Boston: De Gruyter.
Clements, George N. & Engin Sezer. 1982. Vowel and consonant disharmony in Turkish. In Harry van der Hulst & Norval Smith (eds.), The structure of phonological representations, vol. 2, 213–255. Dordrecht: Foris.
Dauer, Rebecca. 1983. Stress-timing and syllable-timing reanalyzed. Journal of Phonetics 11. 51–62.
De Yzaguirre, Lluís. 1995. L'estructura sil·làbica del català central. Barcelona: Universitat de Barcelona dissertation. http://latel.upf.edu/morgana/altres/cibres/tesi_lym/estrucsil.htm (accessed 25 November 2017).
DiCanio, Christian T. 2008. The phonetics and phonology of San Martín Itunyoso Trique. Berkeley, CA: University of Berkeley dissertation. http://linguistics.berkeley.edu/dissertations/DiCanio_dissertation_2008.pdf (accessed 25 November 2017).
DiCanio, Christian T. 2010. Itunyoso Trique. Journal of the International Phonetic Association 40(2). 227–238.
Dienst, Stefan. 2014. A grammar of Kulina (Mouton Grammar Library 66). Berlin & Boston: Mouton de Gruyter.
Dixon, Robert M. W. & Alexandra Y. Aikhenvald (eds.). 2002. Word: A cross-linguistic typology. Cambridge: Cambridge University Press.
Gavaldà-Ferré, Nùria. 2007. Vowel reduction and Catalan speech rhythm. London: University College London MA thesis.
Gilles, Peter. 2014. Phonological domains in Luxembourgish and their relevance for the phonological system. In Javier Caro Reina & Renata Szczepaniak (eds.), Syllable and word languages (linguae & litterae 40), 279–304. Berlin & Boston: De Gruyter.
Greenberg, Joseph H. 1978. Some generalizations concerning initial and final consonant clusters. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), Universals of human language, vol. 2, 243–279. Stanford: Stanford University Press.
Greenberg, Joseph H., Charles E. Osgood & James J. Jenkins. 1966. Memorandum concerning language universals. In Joseph H. Greenberg (ed.), Universals of language, 2nd edn., xv–xxvii. Cambridge, MA: MIT Press.

Hall, T. Alan. 1999. The phonological word: A review. In T. Alan Hall & Ursula Kleinheinz (eds.), Studies on the phonological word (Current Issues in Linguistic Theory 174), 1–22. Amsterdam & Philadelphia: John Benjamins.
Hall, T. Alan. 2011. Phonologie: Eine Einführung, 2nd edn. Berlin & New York: Walter de Gruyter.
Hall, T. Alan & Ursula Kleinheinz (eds.). 1999. Studies on the phonological word (Current Issues in Linguistic Theory 174). Amsterdam & Philadelphia: John Benjamins.
Hannahs, S. J. 2013. The phonology of Welsh. Oxford: Oxford University Press.
Hunziker, Daniel A., Elisabeth Hunziker & Helen Eaton. 2008. A description of the phonology of the Sandawe language (SIL Electronic Working Papers 2008–004). http://sil.org/resources/publications/entry/7875 (accessed 25 November 2017).
Hyman, Larry M. 2008. Directional asymmetries in the morphology and phonology of words, with special reference to Bantu. Linguistics 46(2). 309–350.
Josephs, Lewis S. 1975. Palauan reference grammar. Honolulu: The University Press of Hawaii.
Julià, Joan. 2002. Els sons del català. In Joan Solà, Maria-Rosa Lloret, Joan Mascaró & Manuel Pérez Saldanya (eds.), Gramàtica del català contemporani, vol. 1, 37–87. Barcelona: Empúries.
Kabak, Barış. 2011. Turkish vowel harmony. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), The Blackwell companion to phonology, vol. 5, 2831–2854. Malden, MA: Wiley-Blackwell.
Kabak, Barış. 2014. Pervasive syllables and phonological unity in words. In Javier Caro Reina & Renata Szczepaniak (eds.), Syllable and word languages (linguae & litterae 40), 112–139. Berlin & Boston: De Gruyter.
Kohler, Klaus J. 1994. Glottal stops and glottalization in German. Data and theory of connected speech processes. Phonetica 51. 38–51.
Kohler, Klaus J. 1995. Einführung in die Phonetik des Deutschen, 2nd edn. (Grundlagen der Germanistik 20). Berlin: Erich Schmidt.
Kouwenberg, Silvia. 2010. Creole studies and linguistic typology. Part 2. Journal of Pidgin and Creole Languages 25(2). 359–380.
Kümmel, Martin. 2014. Syllable- and word-related developments in earlier Indo-Iranian. In Javier Caro Reina & Renata Szczepaniak (eds.), Syllable and word languages (linguae & litterae 40), 204–221. Berlin & Boston: De Gruyter.
Levelt, Clara C. & Ruben van de Vijver. 2004. Syllable types in cross-linguistic and developmental grammars. In René Kager, Joe Pater & Wim Zonnefeld (eds.), Constraints in phonological acquisition, 204–218. Cambridge: Cambridge University Press.
Lewis, Geoffrey L. 1967. Turkish grammar. Oxford: Clarendon.
Lloret, Maria-Rosa. 2002. Estructura sil·làbica. In Joan Solà, Maria-Rosa Lloret, Joan Mascaró & Manuel Pérez Saldanya (eds.), Gramàtica del català contemporani, vol. 1, 195–249. Barcelona: Empúries.
Lloret, Maria-Rosa & Jesús Jiménez. 2008. Marcatge posicional i prominència en el vocalisme àton. Caplletra 45. 55–91.
Lloret, Maria-Rosa & Jesús Jiménez. 2009. Phonological variation in voicing across word boundaries. Paper presented to Old World Conference in Phonology 6, University of Edinburgh. http://uv.es/foncat/cat/Treballs/26.Lloret-Jimenez.pdf (accessed 25 November 2017).
Low, Ee Ling, Esther Grabe & Francis Nolan. 2000. Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. Language and Speech 43(4). 377–401.

Maddieson, Ian. 2013a. Syllable structure. In Matthew S. Dryer & Martin Haspelmath (eds.), The world atlas of language structures online. Munich: Max Planck Digital Library. http://wals.info/chapter/12 (accessed 25 November 2017).
Maddieson, Ian. 2013b. Consonant inventories. In Matthew S. Dryer & Martin Haspelmath (eds.), The world atlas of language structures online. Munich: Max Planck Digital Library. http://wals.info/chapter/1 (accessed 25 November 2017).
Maddieson, Ian. 2013c. Vowel quality inventories. In Matthew S. Dryer & Martin Haspelmath (eds.), The world atlas of language structures online. Munich: Max Planck Digital Library. http://wals.info/chapter/2 (accessed 25 November 2017).
Maddieson, Ian. 2013d. Tone. In Matthew S. Dryer & Martin Haspelmath (eds.), The world atlas of language structures online. Munich: Max Planck Digital Library. http://wals.info/chapter/13 (accessed 25 November 2017).
Martínez, Eugenio, Ana María Fernández & Josefina Carrera. 2003. Castilian Spanish. Journal of the International Phonetic Association 33(2). 255–259.
Mascaró, Joan. 2002. El sistèma vocalic: Reducció vocàlica. In Joan Solà, Maria-Rosa Lloret, Joan Mascaró & Manuel Pérez Saldanya (eds.), Gramàtica del català contemporani, vol. 1, 89–123. Barcelona: Empúries.
Moser, Virgil. 1951. Frühneuhochdeutsche Grammatik. III. Band: Lautlehre. 3. Teil: Konsonanten, 2. Hälfte (Schluss) (Germanische Bibliothek Abt. 1. Elementar- und Handbücher Reihe 1. Grammatiken 17). Heidelberg: Winter.
Nespor, Marina, Mohinish Shukla & Jacques Mehler. 2011. Stress-timed vs. syllable-timed languages. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), The Blackwell companion to phonology, vol. 2, 1147–1159. Malden, MA: Wiley-Blackwell.
Nespor, Marina & Irene Vogel. 2007. Prosodic phonology. With a new foreword, 2nd edn. (Studies in Generative Grammar 28). Dordrecht: Foris.
Parker, Steve. 2011. Sonority. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), The Blackwell companion to phonology, vol. 2, 1160–1184. Malden, MA: Wiley-Blackwell.
Parker, Steve. 2012. Sonority distance vs. sonority dispersion – a typological survey. In Steve Parker (ed.), The sonority controversy (Phonology and Phonetics 18), 101–166. Berlin & Boston: De Gruyter.
Paul, Hermann. 1998. Mittelhochdeutsche Grammatik, 24th edn. Tübingen: Niemeyer.
Quilis, Antonio. 1999. Tratado de fonología y fonética españolas, 2nd edn. Madrid: Gredos.
Recasens, Daniel. 1996. Fonètica descriptiva del català. Assaig de caracterització de la pronúncia del vocalisme i consonantisme del català al segle XX, 2nd edn. (Biblioteca filològica 21). Barcelona: Institut d'Estudis Catalans.
Restle, David & Theo Vennemann. 2001. Silbenstruktur. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), Language typology and language universals, vol. 2 (Handbücher zur Sprach- und Kommunikationswissenschaft 20), 1310–1336. Berlin & New York: Walter de Gruyter.
Revithiadou, Anthi. 2011. The phonological word. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), The Blackwell companion to phonology, vol. 2, 1204–1227. Malden, MA: Wiley-Blackwell.
Riad, Tomas. 2014. The phonology of Swedish. Oxford: Oxford University Press.
Szczepaniak, Renata. 2007. Der phonologisch-typologische Wandel des Deutschen von einer Silben- zu einer Wortsprache (Studia Linguistica Germanica 85). Berlin & New York: De Gruyter.

Szczepaniak, Renata. 2009. Wortsprachliches Deutsch und silbensprachliches Spanisch. Ein phonologisch-typologischer Vergleich. Estudios filológicos alemanes 17. 251–267.
Szczepaniak, Renata. 2014. Vowel and consonant epentheses in the history of German from the typological perspective of syllable and word languages. In Javier Caro Reina & Renata Szczepaniak (eds.), Syllable and word languages (linguae & litterae 40), 160–180. Berlin & Boston: De Gruyter.
Traill, Anthony. 1985. Phonetic and phonological studies of !XÓÕ Bushman (Quellen zur Khoisan-Forschung 1). Hamburg: Buske.
Trubetzkoy, Nikolaj S. 1971. Grundzüge der Phonologie. 5th edn. Göttingen: Vandenhoeck & Ruprecht.
Vennemann, Theo. 1988. Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin. Berlin: Mouton de Gruyter.
Veny, Joan. 1991. Katalanisch: Areallinguistik. In Gunter Holtus, Michael Metzeltin & Christian Schmitt (eds.), Französisch, Okzitanisch, Katalanisch, vol. 2 (Lexikon der Romanistischen Linguistik 5), 243–261. Tübingen: Niemeyer.
Veny, Joan. 2002. Els parlars catalans. Síntesi de dialectologia, 13th edn. Mallorca: Moll.
Wheeler, Max. 2005. The phonology of Catalan. Oxford: Oxford University Press.
Wiese, Richard. 1996. The phonology of German. Oxford: Oxford University Press.
Zimmer, Karl & Orhan Orgun. 1992. Turkish. Journal of the International Phonetic Association 22(1–2). 43–45.

Ulrike Domahs, Frank Domahs, and Christina Kauschke

The morphology-prosody interface in typically developing and language-impaired populations

Abstract: This chapter's aim is to scrutinize the role of prosody in processing and producing inflected or derived phonological words. On the basis of three phenomena, it will be shown how morphological operations in German like prefixation and suffixation are constrained by prosodic properties of words. A further aim is to investigate whether children and/or adults with language impairments are sensitive to morpho-prosodic aspects of language. In Section 3 of this chapter, existing psycholinguistic evidence on the acquisition and processing of morpho-prosodic regularities in normal and language-impaired populations will be reviewed. In Section 4, we will exemplify experimental approaches to investigate the morphology-prosody interface, presenting two studies that target the acquisition and the processing of prefixation in German participles. We demonstrate that individuals with language impairment in childhood show subtle to moderate difficulties and peculiarities in the production and perception of prosodically well-formed words.

Keywords: language acquisition, specific language impairment, event-related potentials, participle formation, prefixation

1 Introduction

The phonological word and metrical foot are constituents of the prosodic hierarchy (e.g. Nespor and Vogel 2007) in which diverse phonological operations like resyllabification or stress assignment apply. However, these entities are not only relevant for phonological operations, but under certain conditions also for

Ulrike Domahs: Philipps-Universität Marburg, Pilgrimstein 16, D-35032 Marburg, ulrike.domahs@staff.uni-marburg.de
Frank Domahs: Philipps-Universität Marburg, Pilgrimstein 16, D-35032 Marburg, frank.domahs@staff.uni-marburg.de
Christina Kauschke: Philipps-Universität Marburg, Pilgrimstein 16, D-35032 Marburg, kauschke@staff.uni-marburg.de

DOI 10.1515/9783110542899-005

morphological operations. The latter are covered under the notion of "prosodic morphology" (McCarthy 2006), a term for a discipline that studies how morphological and prosodic characteristics of linguistic forms interact. Within prosodic morphology, restrictions on the size and shape of morphologically simple or complex words are defined that result in certain language-specific prosodic templates. In other words, "the wellformedness of complex words depends on prosodic conditions being fulfilled" (Wiese 2000: 85). In German, word stress and the trochaic foot turn out to be crucial factors determining the distribution of affixes for inflection or derivation.

During language development, children have to acquire such prosodic constraints for complex words along with the morphological constituents (affixes), rules and functions of the morphological paradigm. Thus, the acquisition of morphological paradigms and the prosodic factors that influence the shape of complex words is a demanding and sometimes lengthy task for the learner. Furthermore, it can be assumed that this interface between prosody and morphology might be particularly challenging for individuals with difficulties in language acquisition or processing.

The aim of the present article is to address the two following questions regarding the role of prosodic constraints for the acquisition and processing of selected morphological operations:

– How do typically developing children and children with language impairments master morpho-prosodic constraints of complex words? Do they produce word forms that violate such constraints?
– Are children and/or adults with language impairments less sensitive to the prosodic structure of complex words?

These questions will be addressed in two studies on German participles: a study on children's language production and one on language comprehension in adults.

In the following section, we will introduce in more detail the notion of prosodic morphology and describe three phenomena found in German morphology that are constrained by prosodic properties: Deadjectival derivation, plural formation, and participle formation. The third section reviews previous studies investigating morpho-prosodic processing and existing psycholinguistic evidence on the acquisition and the processing of morpho-prosodic regularities in normal and language-impaired populations. In Section 4, the main section of this article, two different experimental studies are presented that scrutinized effects of prosodic constraints on German participle prefixation. The first study investigates the production of participles in children with and without specific language impairment, while the second study examines the perception of correct and incorrect participle forms in adults with and without a history of language impairment by means of event-related potentials (ERPs). In the final section, we will discuss the relevance of our findings for our understanding of prosodic effects on the processing of morphologically complex words and potential difficulties that may arise in clinical populations.

2 Prosodic Morphology in German

Typological analyses show that certain types of morphological operations can best be formalized when considering the shape of word-prosodic constituents as a prerequisite for the application of inflectional or derivational operations. McCarthy and Prince (e.g., 1994) introduced the scientific field of "Prosodic Morphology" when they found that specific morphological operations obey a particular prosodic template like a foot or a minimal word. Size and form of morphological entities can thus be determined by prosodic properties.

For German, some morphological operations have been suggested to rely on the trochaic pattern of the base form or to result in a certain prosodic template. In the following, this will be demonstrated on the basis of three German phenomena in inflection and derivation: deadjectival nominalizations expressed by the suffixes -*heit* and -*keit*, plural formation, and participle formation.

2.1 Deadjectival nominalization

The suffixes -*heit* and -*keit* both derive abstract nouns from adjectives denoting properties (comparable to the English suffix -*ness*). For both affixes, there are clear prosodic constraints on the final part of their respective bases, but the conditions for the two suffixes vary. For this reason, they are regarded as allomorphs of one single morpheme (see Fleischer and Barz 1995; Wiese 2000). Basically, -*heit* requires base forms that end in a stressed syllable (e.g. *beliebt*$_A$ 'popular') or that are monosyllabic (e.g. *schön*$_A$ 'beautiful'), whereas -*keit* is restricted to base forms that end in an unstressed syllable (e.g. *einsam*$_A$ 'lonely'). In other words, the two affixes are in a complementary distribution to each other, depending on specific prosodic conditions.

2.2 Plural formation

The German plural system involves several more or less predictable inflectional endings, but one predominant prosodic pattern: plural words typically end in a bisyllabic trochee with a reduced final syllable (Eisenberg 2006; Wegener 2002; Wiese 2000). Monosyllabic or finally stressed stems like *Stein* ('stone'), *Hemd* ('shirt'), *Salát* ('salad'), or *Pirát* ('pirate') are suffixed by one of the two syllabic endings *-e* or *-en* (*Steine*, *Hemden*, *Salate*, *Piraten*), while stems ending in a pseudo-suffix like *Wagen* ('car/carriage'), *Muskel* ('muscle'), *Blume* ('flower'), or *Vogel* ('bird') remain either unsuffixed (so-called zero plural; e.g. $Wagen_{Pl}$), take the non-syllabic *-n* suffix (e.g. *Muskeln*, *Blumen*), or are marked for plural by modification of the stem vowel (umlaut; e.g. $\mathrm{Vogel}_{\mathrm{Sg}}$ – $Vögel_{\mathrm{PL}}$), thus preserving their trochaic pattern. Apart from a set of words that is suffixed by the ending *-s* and remains either monosyllabic (e.g. *Tipps*) or does not end in a reduced syllable (e.g. *Omas*), the prosodic condition is met in most plural forms. An examination of a patient with primary progressive aphasia has demonstrated that the prosodic template is a relevant part of adult speakers' representations of grammatical number (Domahs et al. 2017). Even child productions of plural forms obey this prosodic restriction, as will be reported on in Section 3 in more detail.

2.3 Participle formation

To some extent, the analysis of past participles is complementary to those of plural formation: while the optimal prosodic template of German noun plurals is a strong-weak pattern, past participles require the opposite pattern, i.e. they have to begin with an unstressed syllable, followed by a stressed one. While plural marking requires operations at the right edge of prosodic words, the paradigm of participle formation is a case in which the output of the morphological operation is marked with respect to the left edge of prosodic words (in addition to suffixation).

Participle formation in German has often been described as a phenomenon that requires the co-occurrence of prefixation and suffixation. A suffix is always added (either *-t* or *-en*) to mark a participle. In addition, many, but not all participles are marked by the prefix *ge-*. The German *ge*-prefixation in participle formation has been classified as another example of foot-related morphological processes, where the *ge*-prefix attaches to verb stems that begin with a strong (stressed) syllable ($\{[{}^{\text{ˈ}}kauf\text{-}]_{\mathrm{F}}\}_{\omega}$– $\{ge\text{-}[{}^{\text{ˈ}}kauft]_{\mathrm{F}}\}_{\omega}$), but not to stems with a weak initial syllable ($\{ra[sier\text{-}]_{\mathrm{F}}\}_{\omega}$– $*\{ge\text{-}ra[siert]_{\mathrm{F}}\}_{\omega}$) (Geilfuß-Wolfgang 1998; Wiese 2000). Prefixation is therefore not obligatory to express the grammatical function of participles, and has been suggested to serve the requirement that participles have to start prosodically weak. Conditions on the appearance of the prefix *ge-* are phonologically triggered in the sense that "verbs [...] obligatorily taking the prefix [...] are either monosyllabic or, if not monosyllabic, show a stress pattern such that the first syllable following *ge-* bears the primary stress" (Wiese 2000: 90). Since German verbal stems are often stressed on the first syllable, the unstressed *ge-* prefix is highly frequent. In contrast, participles with primary stress on some other syllable in the word cannot be prefixed by *ge-*. In such cases, prefixation of *ge-* would lead to two adjacent unstressed syllables (i.e., stress lapse), which is prosodically disfavoured. Wiese (2000) concludes that word stress is the crucial factor determining the distribution of the prefix.

For all three operations – deadjectival nominalization, plural formation, and participle formation – production and/or perception experiments were performed to test whether children are aware of the relevant prosodic restrictions and obey them in the formation of grammatical words. In the following section, we briefly summarize previous studies on the distribution of the suffixes -*heit* and -*keit* (Janssen, Wiese, and Schlesewsky 2006; Domahs et al. 2013) and on plural formation (Kauschke, Renner, and Domahs 2013; Kauschke, Kurth, and Domahs 2011). Afterwards, two new studies on participle formation are presented in more detail.

3 Prosodic Morphology in Language Acquisition and Processing

Prosody – in particular the dominant stress pattern of words – forms the essential cue for very early language perception in infants, and it has been suggested that prosodic cues bootstrap the segmentation and recognition of word units in many languages (e.g. Höhle et al. 2009; for other aspects of speech segmentation see Boll-Avetisyan 2018, this volume). Later in the stage of first word production, it can be observed that the structure of words is restricted to a particular prosodic pattern. For example, German children prefer bisyllabic trochaic patterns when trying to produce multisyllabic target words (see Grimm 2010, for simple words and compounds). In addition, a few results reported in the literature suggest that prosodic information may also play an important role for the acquisition of morphologically complex (inflected or derived) word forms. We will report studies that investigate the acquisition of specific morphological paradigms in German and ask whether prosodic constraints are vulnerable in children with specific language impairment.

Specific Language Impairment (SLI) is an impairment of language comprehension, language production, or both in the absence of hearing impairment, general developmental delay, neurological disorder or autism (Schwartz 2009). Obvious language problems predominate in the pre-school and early school years, but (subtle) limitations in the perception or production of complex linguistic forms may persist until adulthood (e.g. Rescorla 2009; see Section 4.2). Deficits in the area of syntax and morphology are well-known and well-described characteristics of SLI, whereas the interface between morphology and prosody has received little attention so far. Schwartz (2009) points out that patterns of morphological deficits seem to reflect the prosodic characteristics of the given language and that the production of inflections may be influenced by the phonological and prosodic structure of words. Similarly, Leonard (2009) claims that studies of children with SLI in Germanic languages should place more emphasis on possible prosodic influences on the use of grammatical morphology.

Only few studies addressed this interface so far. In the following, we will review the scant studies on the acquisition and processing of (partly) prosodically conditioned morphology, namely on German deadjectival nominalization and plural formation in children with and without SLI.

3.1 Deadjectival nominalization

In an elicitation experiment, Domahs et al. (2013) aimed at investigating at what age German children master the prosodic and morphological constraints of the word formation paradigm *-heit*/*-keit*, and whether children with SLI have particular difficulties identifying prosodic cues from the input. As described in the previous section, derived words with *-heit* contain simple bases (morphological cue) with final stress (prosodic cue) and those with *-keit* have complex bases with a weak final syllable. Three groups of typically developing children (four, six and eight years old) and 18 children with SLI (aged 8 to 10 years) were asked to produce *-heit* or *-keit* derivations in a sentence completion task including words and pseudowords. As revealed by the elicitation task, four-year-old children performed at chance level and did not seem to have acquired any of the constraints yet. Six-year-old children performed above chance with those stems that simultaneously fulfil both sets of constraints (morphological as well as prosodic). Thus, they had acquired the basic knowledge relevant for the word formation paradigm *-heit* and *-keit*, but still have to master less clear cases. Eight-year-old children performed nearly adult-like and had also acquired the suffix distribution for those cases violating either prosodic or morphological cues. In contrast, children with SLI did not produce systematic responses that follow prosodic and/or morphological constraints. Thus, the findings support the assumption that children with SLI are less sensitive to prosodic properties of grammatical forms than typically developing peers.

3.2 Plural formation

The acquisition of the German plural system has been investigated extensively (e.g. Clahsen et al. 1992; Szagun 2001; Laaha et al. 2006; Kauschke, Kurth, and Domahs 2011). The findings of our own cross-sectional study on noun plural formation with 60 monolingual, typically developing children between three and six years of age (Kauschke, Kurth, and Domahs 2011) converged with findings from other studies with respect to the developmental sequence of the various plural markers: the suffixes *-e* and *-e* plus vowel change (umlaut) were mastered early and easily, whereas *-s* and *-er* plurals caused more problems. Plural forms that require umlaut but no overt suffix were particularly challenging even for older children. Performance in an elicitation task with 44 word stimuli remained on a sub-ceiling plateau between age 3;0 to 4;11; afterwards there was a marked improvement. Between 5 and 6 years of age, 80% of the children's responses were correct, suggesting that full mastery of the German plural system is achieved relatively late in first language acquisition. Among the developmental stages of individual suffixes, the study revealed further interesting correlations between the choice of a suffix and the prosodic shape of the stem: Typically developing children preferably overapplied the suffix *-e* to monosyllabic stems (e.g. **Bäre* instead of *Bären* 'bears') and added *-s* to stems already ending in a trochee (e.g. **Vaters* instead of *Väter* 'fathers'). Thereby, the children showed a tendency to meet the prosodic constraints for German noun plurals. To produce a canonical trochee with a reduced final syllable, they choose the most unmarked way by adding a final schwa (*-e*). Although the children produced morphological errors in plural markings, the prosodic structures of pluralized nouns obeyed prosodic well-formedness. Eight children with SLI performing the same task produced significantly more errors than typically developing children of the same age and performed at the level of younger typically developing children.

Given that the word material in Kauschke, Kurth, and Domahs' (2011) study was not sufficiently designed to test the sensitivity to the prosodic restriction on plural formation in children with and without SLI, a follow-up study tested responses to a larger set of mono- and bisyllabic words and pseudowords. In this elicitation study (Kauschke, Renner, and Domahs 2013), the plural formation ability of 14 German-speaking children with SLI (mean age 7;5 years) was compared to those of age-matched controls and a younger control group (mean age 5;7 years). In line with the study mentioned above (Kauschke, Kurth, and Domahs 2011), children with SLI fell behind the group of age-matched children in their overall performance and even performed significantly worse than the younger control group. More importantly, detailed error analyses revealed that children with SLI produced more forms that did not meet the optimal prosodic shape of a noun plural. Whereas typically developing children predominantly omitted plural suffixes when the word stem was already trochaic (e.g. **Leiter* instead of *Leitern* ‘ladders’), the children with SLI produced a higher proportion of omissions in the context of stems with final stress (e.g. **Pirat* instead of *Piraten* ‘pirates’). As a consequence, fewer of their plural forms were trochaic and thus prosodically well-formed. Beyond the fact that children with SLI have deficits in plural marking, these findings suggest that they also show reduced sensitivity to prosodic aspects of plural formation.

In light of the results obtained from previous studies, we can conclude that children with SLI are less sensitive to morpho-prosodic requirements of complex words than typically developing children. Our aim is now to extend the research focus, examining the acquisition of prosodically modified morphological operations in the paradigm of German participle formation. The German participle formation, in particular the prefixation, is an ideal test case for experimental studies on the prosody-morphology-interface due to the fact that the *ge*-prefixation satisfies prosodic requirements of the grammatical word (i.e., participle), which can be manipulated independently from the suffix.

In the following section, two studies are presented that scrutinized prefixation in participles systematically. The aim of the first study was to investigate how typically developing children master the prosodic constraints on participle formation and whether these constraints are vulnerable in children with specific language impairment. In the second study, event-related potentials (ERPs) were measured during the processing of words in which the prosodic conditions for prefixation in participles were violated (e.g. (*hat*) **ge+studiert*).

4 Studies on German Participle Prefixation

4.1 Study 1: Participle formation in children with and without language impairment

The acquisition of past participles in German is a complex task since children have to learn

(a) the distribution of the two suffixes *-t* and *-(e)n*,
(b) the patterns of stem vowel modification (ablaut) in strong verbs, and
(c) the distribution of the prefix *ge-*.

While the morphological aspects of regular and irregular suffixation have been investigated exhaustively in previous research (Clahsen and Rothweiler 1993; Clahsen et al. 2014; Elsen 1998; Lindner 1998; Szagun 2011; Wittek and Tomasello 2002), it is still unclear how children master the prosodically determined prefixation rule, namely that prefixation of *ge-* depends on the stress pattern of the verbal stem. According to Szagun's (2011) data, prefix omissions are mainly found in early stages of language development, i.e. in the second and third year of life. This early phase of prefix omissions is then followed by a decrease over time. Previous studies are also inconsistent with respect to the question whether problems of children with SLI extend to participle formation. Some studies suggest that past participle formation, both in English and German, is not affected (Redmond 2003; Clahsen et al. 2014), whereas others claim that problems with past participles are part of the morphological symptoms associated with SLI (Leonard et al. 2003).

To test the sensitivity to the prosodic conditions for prefixation, we conducted an elicitation study with typically developing children and children with SLI (see Kauschke, Renner, and Domahs 2017, for more details). In other words, we asked the question whether children with and without SLI adhere to the prosodic constraints that *ge-* is required in participles in order to obey the prosodic template that participles begin with an unstressed syllable or not?

4.1.1 Methods

4.1.1.1 Participants

In total, 41 children from monolingual German homes participated in the study. The sample was identical to the one described in Kauschke, Renner, and Domahs (2013, see above, and in Kauschke, Renner, and Domahs 2017). Nonverbal intelligence was assessed with a standardized test (CPM, Bulheller and Häcker 2006), and only those children were included who scored in the normal range. With respect to language abilities, two age-appropriate standardised tests were applied in order to assess vocabulary production (for older children: WWT, Glück 2007, for younger children: AWST-R, Kiese-Himmel 2005) as well as grammatical skills (sentence comprehension: TROG-D, Fox 2006). Details on the test results are reported in Kauschke, Renner, and Domahs (2013, 2017). Three groups of children were formed:

– 14 Children with SLI (six girls, eight boys) with a mean age of 7;5 years. All children were diagnosed as language impaired by clinicians and attended a school for special education. The children's medical histories assured normal hearing, normal neurological status, and no pervasive disorders. The inclusion criterion for the children with SLI was below-average performance in at least one of the two language tests. In addition to their lexical and/or grammatical deficits, seven of the 14 children also showed symptoms of speech sound disorders.
– 14 typically developing (TD) children (nine girls, five boys) of the same chronological age as the children with SLI (mean age 7;3).
– 13 younger typically developing children (six girls, seven boys, mean age 5;7) whose level of lexical production was comparable to the children with SLI. Each of the typically developing children in both groups had to score in the normal range in both language tests.

4.1.1.2 Material and Procedure

The complete material consisted of 60 words and 20 pseudowords, but we will focus on the word set here. The 60 word stimuli comprised regular German verbs that form their past participle with the suffix *-t* without a vowel change. 30 of these verbs started with a stressed syllable and thus required the prefix *ge-* (e.g. *flöten – geflötet* 'play/played the flute'), while the other 30 verbs started with an unstressed syllable and thus did not allow the prefixation of *ge-* (e.g. *trompéten – trompétet* 'play/played the trumpet', but: **getrompétet*). Ten of these verbs contained an unstressed prefix (*ver-*, *be-*, *zer- er-*, *ent-*) in their present tense form and did not allow for the prefixation with *ge-* (e.g. *bezáhlen – bezahlt*, 'pay – payed', but: **gebezáhlt*). The sets of verbs with and without *ge-* prefixation were matched for frequency and age of acquisition.

The method used was a cross-modal sentence completion task. Each trial started with the presentation of a short video sequence showing familiar actions. Simultaneously, the critical verb was presented auditorily twice embedded in a carrier sentence (e.g. 'Look here, the people are working. They are working.’). The auditory stimuli were produced by a trained female speaker and presented on a laptop computer. The child’s task was to complete a subsequent question (‘What have they done? They have…’, also presented auditorily) by producing a past participle (‘worked’). The children’s responses were recorded and transcribed for further analysis.

4.1.1.3 Data analysis

In order to calculate accuracy, responses were coded as correct or incorrect. A response was counted as morphologically correct when both the appropriate prefix (if necessary) and the appropriate suffix of the target were produced correctly. Phonological errors were ignored in the coding of morphological accuracy as long as the inflectional affixes could be clearly identified. Incorrect responses were then classified according to the type of error:

– Omissions, additions or substitutions of the prefix
– Omissions, additions or substitutions of the suffix
– Responses without participle marking
– Nonword responses
– Responses with a word different from the target word
– Refusals.

We were especially interested in inadequate additions or omissions of the prefix *ge-* since these errors can be considered as violations of the prosodic constraint for participles.

4.1.2 Results

With respect to general accuracy of participle formation, results show that both groups of typically developing children produced more correct responses than the children with SLI. Figure 1 illustrates that the age-matched TD children marked the participles correctly in 95.8% (SD 4.4) of all responses, compared to 87.5% (SD 15.9) for the younger TD children and only 64.0% (SD 13.6) for the children with SLI. A Kruskal-Wallis one-way analysis of variance revealed a highly significant group effect ($p < .001$). Post-hoc tests clarified that both groups of TD children produced significantly more correct responses than the children with SLI (SLI vs. older TD group: $U = 4.5$, $p < .001$; SLI versus younger TD group: $U = 21$, $p < .01$). The difference between the older and the younger control group was marginally significant ($p < .07$).

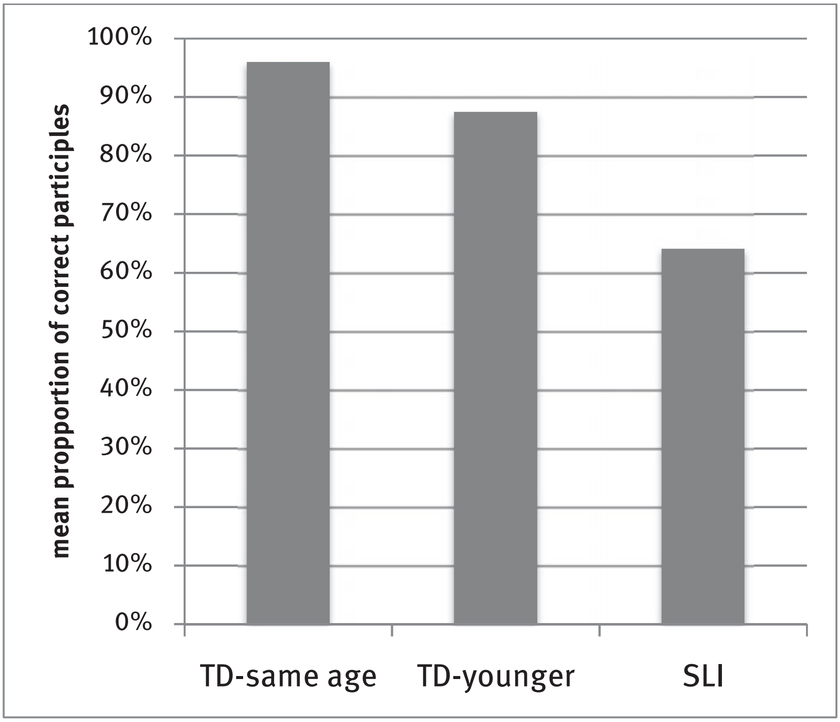


Fig. 1: Mean accuracy for participle formation

In a detailed error analysis, we calculated the errors on the basis of all analysable responses (i.e., correct plus incorrect responses, excluding non-analysable responses like refusals, other words, and nonword responses). Again, the Kruskal-Wallis-Test revealed a significant group effect ($p < .01$) for the responses without participle markings. Pairwise comparisons showed that the children with SLI differed significantly from the older ($U = -2.94$, $p < 0.01$) and from the younger TD group ($U = -2.26$ $p < 0.05$). Obviously, children with SLI avoided participle markings and instead repeated the infinitive or the present tense form more often (6.25%) than TD children (0.51% age-matched TD group/ 2.88% younger TD group).

Regarding errors that clearly indicate violations of the prosodic requirements for past participles, it turned out that these error types occurred very rarely. Therefore, no statistical analyses were conducted. One or two occasional prefix omissions were observed in individual children of all groups. However, one single child with SLI stood out who produced a higher number of prefix omissions (seven omissions, e.g. **schält* instead of *geschält* ‘peeled’).

The number of inadequate prefix additions was also extremely low. None of the TD children added the prefix *ge-* to a verb stem for which prefixation was blocked. However, three *ge*-additions in verbs with initial weak syllables were observed in the responses of two children from the SLI group (e.g. **gerasiert* instead of *rasiert* ‘shaved’).

Additional error analyses revealed that children with SLI exhibit problems with the adequate choice of the prefix, resulting in a significantly higher proportion of prefix substitutions (e.g. *verziffert* instead of *entziffert* ‘deciphered’). However, this error type reflects morphological, rather than morpho-prosodic difficulties and will therefore not be considered in the context of this chapter (see Kauschke, Renner, and Domahs 2017, for more details).

4.1.3 Intermediate Discussion

The elicitation study presented here investigated potential deficits in participle formation in children with SLI with a special emphasis on the application of the prefix *ge-*, which is prosodically determined. The results for general accuracy point to advanced mastery of participle formation at pre-school and beginning school age in typically developing children and – at the same time – to substantial deficits in participle formation in children with SLI. This finding – combined with the fact that children with SLI avoided participle markings – strongly supports the assumption that problems with participle formation are part of the morphological symptoms that characterize SLI in German.

With respect to the morpho-prosodic aspects of participle formation, TD children showed a high degree of sensitivity, consistent with our findings on plural formation. Even the children with SLI mainly adhered to the morpho-prosodic regularities of prefixation, suggesting that prosodic constraints of prefix application in participle formation are relatively robust. However, some prosodic violations were observed in individual children. In particular, one child from the SLI group showed a strong tendency to omit prefixes. This finding is in line with a finding reported by Clahsen and Rothweiler (1993), who also described a child with a high proportion of prefix omissions. The other type of prosodic violation, i.e. prefix addition, was extremely rare. However, such errors occurred exclusively in children with SLI. Taken together, even though prosodic errors may be observed in single cases, the results do not support a comprehensive morpho-prosodic deficit in SLI.

Although prosodically illegal prefix additions like **gerasiert* occur only rarely in responses produced by individual children with SLI, we were interested in the question how these kinds of violations are processed in on-line perception and whether sensitivity to these violations depends on language proficiency. Therefore, we conducted a second study on the processing of German participles, this time in adults with and without a history of language impairment. The objective is guided by the rationale that this methodology may be more sensitive to subtle peculiarities in prosodic processing of *ge*- than production methods.

4.2 Event-related potentials on morpho-prosodic violations of German participle-prefixation

One crucial finding of the acquisition study was that even children with language impairment obey the prosodic restriction on *ge*-prefixation in German participles in the majority of their productions. Whether the few cases of inadequate omissions and additions of the *ge*-prefix might hint to more profound difficulties in this area cannot be answered by a production task alone. Therefore, an ERP study was performed which investigated how the brain of participants with typical and atypical language development responded to correctly and incorrectly prefixed participles. This study was performed with adults who were diagnosed with speech and language impairment in childhood and with an adult control group.

Previous investigations showed that atypical language processes persist in adults with a history of childhood language impairment. For instance, Rescorla (2002, 2005, 2009) found that seventeen-year-old adolescents scored in the normal range in most language tests but performed significantly worse in lexical and grammatical tasks compared to age-matched peers who developed typically. More specifically, the adolescents with a history of language impairment had more difficulties in grammaticality judgement tasks, in correcting ungrammatical sentences, and in sentence comprehension, and were limited in their verbal working memory. Accordingly, Rescorla suggests "subclinical weakness in the component skills that serve language" (Rescorla 2009: 28). Comparable findings are reported by Miller, Leonard, and Finneran (2008), Poll, Betz, and Miller (2010), and Whitehouse et al. (2009).

In an EEG experiment, Fonteneau and van der Lely (2008) observed a lack of an early left-anterior negativity (ELAN) in morpho-syntactic violations in adolescents with grammatical SLI (G-SLI). G-SLI refers to a subgroup of SLI with a persistent and selective deficit in syntax and morphology. The ELAN typically occurs as reflection of higher costs in morpho-syntactic processing (Friederici 2002). Fonteneau and van der Lely (2008) found different neuronal circuitry involved in the processing of syntactic structures compared to typically developing adolescents (N400 instead of an ELAN), suggesting that adolescents with G-SLI compensate for their limitation in syntactic processing. In particular, Fonteneau and van der Lely (2008) propose that early and automatic domain-specific components are affected by the impairment, while later domain-general components are preserved. Converging evidence is also reported by Sabisch et al. (2009) for German children with SLI.

Based on such results the main objective of the present study was to measure ERP components that reflect the processing of correctly and incorrectly prefixed participles in adults with and without a history of language impairment. Although we found that the prosodic requirements in participle formation are mostly fulfilled in production tasks, it is worthwhile to investigate the perceptual processing of violations of such requirements. A further important question here is which processing steps are affected by prefixation violations. Do incorrect prefixations of *ge-* lead to enhanced costs in prosodic, morpho-syntactic or lexical processing? In the first case, we are expecting either an early bilateral anterior negativity (Sabisch et al. 2009; Schmidt-Kassow and Kotz 2009) or an early positivity (P200; Friedrich, Alter, and Kotz 2001; Shahin et al. 2005; Henrich et al. 2014). If morpho-syntactic processing is affected, an altered left anterior negativity (LAN) is expected (Weyerts et al. 1997; Gross et al. 1998; Gunter, Friederici, and Schriefers 2000). And in the case of additional lexical processing costs, a modified N400 effect should occur (Weyerts et al. 1997; Janssen, Wiese, and Schlesewsky 2006).

4.2.1 Methods

4.2.1.1 Participants

All participants were students or academic staff members of the University of Marburg, Germany. Twenty native speakers of German (12 females) participated in the control group. All control participants were right-handed, without a history of language, hearing, or other disorders. Their mean age was 25 years (ranging from 21 to 49 years).

Eighteen right-handed adults (10 females) participated in the group of adults with a history of language impairment (denoted as exLI-group in Section 4.2.2 “Results”). All participants of this group had normal hearing and normal or corrected to normal vision. Their mean age was 26 years (ranging from 19 to

49 years) and they reported to have been treated for speech and language impairments in childhood (mean treatment duration: 2;2 years). The type of childhood language disorders was assessed by a questionnaire adopted from Arkkila (2009), according to which 13 participants reported developmental dyslexia, ten speech sound disorders, five a delay in language development, two syntactic-morphological problems, and one difficulties at the level of text processing (the questionnaire allows to indicate multiple symptoms).

Each participant was assigned to one of the two groups on the basis of the questionnaire.

4.2.1.2 Material

In order to test the processing of correct and incorrect participle prefixation, we selected 15 regular German verbs that consisted of two stem syllables with final word stress (e.g. *studier-* 'to study') that are not prefixed by *ge-*. Each inflected verb was embedded in a sentence (see Example (1)) with the auxiliary verb *haben* ('to have'), and each verb was transitive, selecting an accusative object. The verbs were further controlled for frequency based on the SUBTLEX database (Brysbaert et al. 2011) and for Age of Acquisition (AoA).

(1) *Die Forscher haben die Karte studiert*.
'The scientists have studied the map.'

Each sentence was recorded from a female trained German speaker, realized with either the correct participle form without prefix (*studiert*) or with an incorrect verb form that involved inadequate prefixation(**ge-studiert*).

4.2.1.3 Procedure

Each participant was seated in a sound-attenuated cabin and listened to the auditorily presented sentences that ended either in a correct or incorrect participle. Participants listened passively, but were instructed to memorize the verbs that were included in the sentences and to verify after each block of trials whether a given verb had been presented in the block or not. Each trial started with a fixation cross at the centre of the screen, which should be fixated while listening to the auditory sentences to prevent participants from moving their eyes. The inter-trial interval lasted 2 seconds in which participants were allowed to rest their eyes and move their limbs. Each version of a sentence was presented twice, resulting in a total of 60 items (plus 300 fillers) that were presented in 30 blocks.

The EEG data were recorded by means of 32 AgAgCL electrodes with C2 serving as ground electrode. The reference electrode was placed at the left mastoid and the EEG was re-referenced offline over both mastoids. In order to control for eye-movements, the EOG recorded vertical and horizontal movements with electrodes placed below and above the participant's left eye and at the outer canthus of both eyes. Impedances were kept below 5 kΩ. EEG and EOG were recorded with a Brain-Amp amplifier (Brain Products, Germany) with a sampling rate of 500 Hz and filtered offline with a bandpass filter from 0.16 to 30 Hz.

4.2.1.4 Analyses

ERPs were computed for each participant, condition, and electrode. Trials with eye-movement artefacts were removed from the data (11.6%). Averages were calculated starting at the onset of the critical verb extending to 1200 ms thereafter with a pre-stimulus baseline of 200 ms. In repeated measures ANOVAs, we compared the mean voltages measured for the incorrect condition and the correct condition. Time windows for statistical analyses were selected by visual inspection. For each group, we calculated mean-voltage differences between 60 to 230 ms post-onset, 380 to 600 ms, and 930 to 1200 ms over the factors CORRECTNESS (correct vs. incorrect) and REGION (left anterior: F3, FC1, FC5; right anterior: F4, FC2, FC6; left posterior: CP5, CP1, P3; right posterior: CP6, CP2, P4).

4.2.2 Results

Statistical comparisons of mean voltage measures in the time window from 60 to 230 ms revealed a significant difference between the correct and the incorrect condition indicating a positivity effect for the incorrect condition only in the control group, but not in the exLI-group (controls: $F(1; 19) = 6.93$; $p = .016$; $\eta_p^2 = .27$; exLI-group: $F(1; 17) = .99$; $p = .333$; $\eta_p^2 = .06$).

Furthermore, the condition with incorrect prefixation evoked a significant positivity effect between 380 to 600 ms and a significant positivity effect between 930 to 1200 ms. Both effects were found for both groups of participants (negativity between 380 and 600 ms controls: $t(19) = 2.18$; $p = .042$; exLI-group: $t(17) = 3.511$; $p = .003$; positivity between 930 and 1200 ms controls: $F(1; 19) = 18.45$; $p < .001$; $\eta_p^2 = .49$; exLI-group: $F(1; 17) = 9.05$; $p = .008$; $\eta_p^2 = .35$). Figure 2 depicts the grand average curves obtained for the two groups.

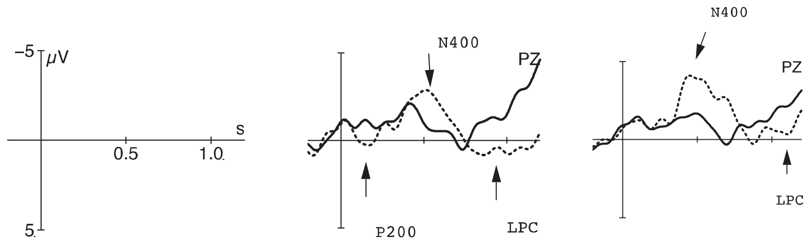


Fig. 2: Grand average curves measured at the central parietal electrode Pz plotted from onset of the correct (solid line) and incorrect (dotted line) participles. The middle panel depicts results obtained from the control group and the right panel from the group with a history of developmental language impairment (exLI).

4.2.3 Intermediate Discussion

Comparisons between correct and incorrect participle forms revealed three different components: an P200 between 60 to 230 ms, an N400 between 380 and 600 ms, and a late positive component (LPC) between 930 and 1200 ms. For the control group, significant P200, N400, and LPC effects were found, for adults with a history of language impairment, an N400 and LPC, but no P200 effect (see Figure 2) were observed.

The occurrence of the three components in the control group suggests different cognitive reflexes to the processing of incorrectly prefixed participles. First, regarding the P200 effect, we follow the interpretation according to which the component indexes the comparison of an (auditory) input with an internal representation or expectation in memory or language context (Evans and Federmeier 2007). The question arises why the morpho-prosodic violation induced such an early effect, occurring before the verb stem was fully encountered. We hypothesize that due to embedding of the target words in sentences that primed participles, the participants expected to hear a participle starting with the prefix *ge-* followed by a stressed syllable. Instead, they encountered *ge-* followed by a weak syllable. The perception of an unexpected sequence of two adjacent weak syllables (i.e., stress lapse) may have led to the early response to deviating material. This explanation converges with other findings: Within a prosodic task, in which the preference for a rhythmically alternating stress pattern was investigated, Henrich et al. (2014) observed an early positivity for a rhythmically deviating strong syllable (after a strong syllable). In addition, further studies support the view that the P200 is sensitive to physical properties like pitch and pitch contours and can be detected even during the processing of initial syllables (Friedrich, Alter, and Kotz 2001; Shahin et al. 2005).

The second component, the negativity between 380 and 600 ms with an emphasis in centro-parietal regions, can be interpreted as an instance of an N400-effect indicating enhanced costs in lexico-semantic integration of a word form that does not exist (**ge-studiert*). This is in line with previous interpretations of such a negativity effect in the literature (for reviews see Friederici 2002; Kutas and Federmeier 2011). It should be noted here that an interpretation as an ELAN might be plausible as well, a component evoked by morpho-syntactic violations with a distribution over left electrode sites (for a review see Friederici 2002). However, the latency and localisation of the component in question speaks against an ELAN: The observed negativity occurred in a later time window than the ELAN (which typically peaks around 150 to 200 ms) and was distributed broadly over both hemispheres with a clear focus on centro-parietal regions which is typical for the N400 component. Consequently, the participle with incorrect prefixation does not violate morpho-syntactic but rather morpho-lexical conditions. It is very likely that a non-existing word form like **ge-studiert* is processed like a pseudoword. Therefore, our suggestion is that forms violating morpho-prosodic conditions produced enhanced costs in lexical integration rather than in morpho-syntactic processing. This is in line with findings reported in Janssen, Wiese, and Schlesewsky (2006).

Finally, the LPC occurring between 930 and 1200 ms with a centro-parietal distribution might reflect the re-analysis of forms encountered that are non-existing, though transparent from their morphological architecture. Given that the participants were not expected to judge on the forms explicitly, it does not seem likely that this late positivity reflects task-related processing steps. Rather, we assume that participants re-constructed the incorrect concatenation of morphemes within the presented participles. This interpretation is consistent with a P600 analysis (for a review see Friederici 2002).

So far, we focused on the brain responses observed in the control group. The group of adults with a history of language impairment (exLI), however, showed an N400 and an LPC effect, but no P200 effect. Thus, early morpho-prosodic processing seems to be affected (lack of P200 effect), while later processes did not deviate (N400, P600). This is in line with findings presented by Fonteneau and van der Lely (2008) and Sabisch et al. (2009) on the processing of morpho-syntax in adolescents with G-SLI, i.e. a lack of an early component. In contrast to these results, the task used in our study did not involve early morpho-syntactic but morpho-prosodic processing. This difference in violation type is reflected in differential ERP signatures (ELAN vs. P200). Thus, our results extend previous findings on morpho-syntactic processing differences in groups with and without language impairment to differences in morpho-prosodic processing. Another difference to previous studies is that the language symptoms of our participants with a history of language impairment were not restricted to grammatical problems that characterize G-SLI but were rather heterogeneous.

Crucially, processing peculiarities in developmental language disorders or childhood dyslexia seem to persist in a subtle way even in young adults. Adults with a history of developmental language disorder or childhood dyslexia showed deviant early phases of morpho-prosodic processing in an online task. This is remarkable given the fact that all participants were educated at an academic level, i.e. verbal skills play an important role in their everyday life. All participants reported that at present they do not experience obvious difficulties with written and spoken language.

5 General Discussion

The main objective of the present article was to investigate how typically developing children and children with language impairments master the acquisition of morpho-prosodic constraints on inflection and word formation. Of particular interest was to test whether morphological operations that are constrained by prosodic requirements are particularly demanding for individuals with SLI and lead, therefore, to prosodic violations in production and/or reduced sensitivity to the prosodic structure of words in language perception.

Previous studies showed that the production of grammatical words in typically developing children is constrained by prosodic requirements (where necessary) and that incorrectly suffixed words rarely violate prosodic conditions. Even in the acquisition of complex word formation paradigms like the German noun-forming paradigm *-heit*/*-keit,* children start by using derived words that obey the prosodic constraint. In contrast to typically developing children, children with specific language impairment showed to be less sensitive to such prosodic aspects of morphology. The production data for plural formation as well as for derivation with *-heit*/*-keit* that were summarized in Section 3 demonstrate that typically developing German children are strikingly sensitive to morpho-prosodic regularities. Mastery of complex morphological paradigms (e.g. the distribution of *-heit* and *-keit* or some aspects of the plural system) may take a long time to develop, but basic requirements for the prosodic wellformedness of words are obeyed from early on. For children with SLI, a mixed picture emerges beyond the fact that they perform significantly worse than age-matched and younger children: Severe violations (such as plurals with two adjacent weak syllables at the right edge, e.g. *Tigere) did not occur. However, the plural studies suggest that children with SLI seem to be less guided by form-based restrictions to produce well-formed trochaic plurals. Despite a certain degree of robustness, SLI seems to be associated with a reduced sensitivity to morpho-prosodic regularities. This in turn may affect successful learning of morphological paradigms.

The two studies on production and perception of *ge*-prefixation presented in Section 4 add to previous findings. Though children with SLI – like typically developing children – turned out to attach the prefix *ge*- correctly to verb stems starting with primary stress, brain responses of adults with a history of developmental language impairment to incorrectly prefixed words differed from those of the control group. In particular, the early P200 component indexing the processing of morpho-prosodic violations was lacking, while later components (N400 and LPC) indicating enhanced costs in lexical processing and re-analysis processes were preserved.

Crucially, the perception study illustrates that certain processing patterns (like the lack of the P200 effect and typical occurrence of the two later components) can only be identified if implicit language processing is investigated by means of highly time-sensitive measures. As a possible explanation for the diverging findings in perception and production, we suggest deviating processing steps in early phases in which speech is categorized as being in agreement with a certain (acoustic) expectation or not. Such early processing deviations can be detected only in perception. In speech production, the output of the formulator (in terms of the speech production model proposed by Levelt, Roelofs, and Meyer 1999) is also subject to internal rehearsal and monitoring processes that might in most cases inhibit incorrectly prefixed participles to occur.

The findings reviewed and presented in this chapter suggest that certain morpho-prosodic abilities seem to be vulnerable in populations with language impairment. In production tasks, this has been shown clearly for plural formation (Kauschke, Renner, and Domahs 2013) and deadjectival nominalization (Domahs et al. 2013), and to a lesser extent for participle formation. In contrast to the plural and -*heit*/-*keit* paradigm, which both allow for exceptions of the prosodic constraint, the morpho-prosodic operation in participle prefixation is highly productive with a clear-cut distribution of the presence or absence of the prefix *ge*-. Our data demonstrate that children with SLI mostly obey these clear regularities of prefix application. The higher predictability might be an important difference between *ge*-prefixation and plural formation or derivation with -*heit* and -*keit* for which less sensitivity in populations with SLI was found.

Obviously, the vulnerability of morpho-prosodic processing depends on the complexity and transparency of the morphological operation to be acquired or processed.

Difficult morpho-prosodic paradigms should therefore be targeted in language intervention. Since intervention traditionally focuses on segmental and/or grammatical aspects of language, a stronger inclusion of prosody-related operations and properties of language in intervention planning is clearly desirable. Hargrove (2013) names lexical stress as one aspect of prosody that is amenable to therapeutic intervention. We propose that language intervention in children with language impairment should draw the child's attention to the prosodic patterns of simple and complex words, in particular to the sequence of weak and strong syllables. When targeting the domain of plural formation, for example, it should be pointed out that German plural forms typically consist of two syllables (strong–weak). In addition, monosyllabic plural forms violating the prosodic requirements of typical plurals (like *Parks* 'parks' or *Loks* 'locomotives') should not be included as target words in the first stages of intervention since these forms offer ambiguous input. Instead, these forms should be introduced later as exceptions. Finally, future research and clinical practice should aim at a systematic evaluation of the effectiveness of prosodic intervention: "Although the research literature on prosodic intervention has expanded in recent years, additional research concerned with the effectiveness of prosodic intervention for all ages and clinical conditions remains the highest priority." (Hargrove 2013: 257).

Acknowledgements: The project was supported by a grant from the Hessian Ministry of Research and the Arts within the LOEWE program "Exploring Fundamental Linguistic Categories" (Landes-Offensive zur Entwicklung Wissenschaftlich-Ökonomischer Exzellenz). The study was approved by the Hessian Ministry of Education and was in accordance with the ethical standards of the World Medical Association (Declaration of Helsinki). The authors thank all children and adults who participated in the studies and Jan Steiner, Lena Renner, and Nina Niggemann for their help in study development, data acquisition and data analyses.

References

Arkkila, Eva. 2009.Specific language impairment in pre-adolescence, adolescence, and adulthood – with special emphasis on health-related quality of life. Dissertation, Department of Phoniatrics, University of Helsinki and Helsinki University Central Hospital, Helsinki.

Boll-Avetisyan, Natalie. 2018. The role of phonological structure in speech segmentation by infants and adults: a review and methodological considerations. In Christiane Ulbrich, Alexander Werth & Richard Wiese (eds.), Empirical approaches to the phonological structure of words (Linguistische Arbeiten 567), 213–240. Berlin & Boston: De Gruyter.
Brysbaert, Marc, Matthias Buchmeier, Markus Conrad, Arthur M. Jacobs, Jens Bölte & Andrea Böhl. 2011. The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. Experimental Psychology 58(5). 412–424.
Bulheller, Stephan & Hartmut Häcker. 2006. Coloured Progressive Matrices: CPM. Manual. Dt. Bearbeitung. Frankfurt: Harcourt Test Services.
Clahsen, Harald & Monika Rothweiler. 1993. Inflectional rules in children's grammars: Evidence from German participles. In Geert Booij & Jaap Marle (eds.), Yearbook of Morphology 1992, 1–34. Dordrecht: Springer Netherlands.
Clahsen, Harald, Monika Rothweiler, Franziska Sterner & Solveig Chilla. 2014. Linguistic markers of specific impairment in bilingual children: The case of verb morphology. Clinical Linguistics & Phonetics 28(9). 709–721.
Clahsen, Harald, Monika Rothweiler, Marcus Woest & Gerry Marcus. 1992. Regular and irregular inflection in the acquisition of German noun plurals. Cognition 45. 225–255.
Domahs, Frank, Lisa Bartha-Doehring, Ulrike Domahs & Margarete Delazer. 2017. Wie muss ein „guter" deutscher Plural klingen? In Nanna Fuhrhop, Renata Szczepaniak & Karsten Schmidt (eds.), Sichtbare und hörbare Morphologie (Reihe Linguistische Arbeiten 565), 205–236. Berlin: De Gruyter.
Domahs, Ulrike, Karin Lohmann, Nicole Moritz & Christina Kauschke. 2013. The acquisition of prosodic constraints on derivational morphology in typically developing children and children with SLI. Clinical Linguistics & Phonetics 27(8). 555–573.
Eisenberg, Peter. 2006. Grundriss der deutschen Grammatik: Das Wort. Stuttgart: Metzler.
Elsen, Hilke. 1998. The acquisition of past participles: One or two mechanisms? In Ray Fabri, Albert Ortmann & Teresa Parodi (eds.), Models of Inflection, 134–151. Tübingen: Niemeyer.
Evans, Karen M. & Kara D. Federmeier. 2007. The memory that's right and the memory that's left: Event-related potentials reveal hemispheric asymmetries in the encoding and retention of verbal information. Neuropsychologia 45. 1777–1790.
Fleischer, Wolfgang & Irmhild Barz. 1995. Wortbildung der deutschen Gegenwartssprache. Tübingen: Niemeyer.
Fonteneau, Elisabeth & Heather K. J. van der Lely. 2008. Electrical brain responses in language-impaired children reveal grammar-specific deficits. PLoS ONE 3(3). e1832. 1–6.
Fox, Annette V. 2006. TROG-D: Test zur Überprüfung des Grammatikverständnisses. Handbuch. Das Gesundheitsforum. Idstein: Schulz-Kirchner Verlag.
Friederici, Angela. 2002. Towards a neural basis of auditory sentence processing. Trends in Cognitive Sciences 6. 78–84.
Friedrich, Claudia K., Kai Alter & Sonja A. Kotz. 2001. An electrophysiological response to different pitch contours in words. NeuroReport 12. 3189–3191.
Geilfuß-Wolfgang, Jochen. 1998. Über die optimale Position von ge-. Linguistische Berichte 176. 581–588.
Glück, Christian W. 2007. Wortschatz- und Wortfindungstest für 6- bis 10-Jährige (WWT 6–10). München: Elsevier.

Grimm, Angela. 2010. The development of early prosodic word structure in child German: Simplex words and compounds. Potsdam. http://opus.kobv.de/ubp/volltexte/2010/4319/ (accessed 14 April 2016).

Gross, Matthias, Tessa Say, Michael Kleingers, Harald Clahsen & Thomas F. Münte. 1998. Human brain potentials to violations in morphologically complex Italian words. Neuroscience Letters 241(2). 83–86.

Gunter, Thomas C., Angela D. Friederici & Herbert Schriefers. 2000. Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. Journal of Cognitive Neuroscience 12. 556–568.

Hargrove, Patricia M. 2013. Pursuing prosodic interventions. Clinical Linguistics & Phonetics 27(8). 574–593, 647–660.

Henrich, Karen, Kai Alter, Richard Wiese & Ulrike Domahs. 2014. The relevance of rhythmical alternation in language processing: An ERP study on English compounds. Brain and Language 136. 19–30.

Höhle, Barbara, Ranka Bijjeljac-Babic, Birgit Herold, Jürgen Weissenborn & Thierry Nazzi. 2009. Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. Infant Behavior and Development 32. 262–274.

Janssen, Ulrike, Richard Wiese & Matthias Schlesewsky. 2006. Electrophysiological responses to violations of morphosyntactic and prosodic features in derived German nouns. Journal of Neurolinguistics 19. 466–482.

Kauschke, Christina, Anna Kurth & Ulrike Domahs. 2011. Acquisition of German Noun Plurals in Typically Developing Children and Children with Specific Language Impairment. Child Development Research. DOI: 10.1155/2011/718925.

Kauschke, Christina, Lena Renner & Ulrike Domahs. 2013. Prosodic constraints on inflected words: An area of difficulty for German-speaking children with specific language impairment? Clinical Linguistics & Phonetics 27(8). 574–593.

Kauschke, Christina, Lena Renner & Ulrike Domahs. 2017. Past Participle Formation in Specific Language Impairment. International Journal of Language and Communication Disorders 52(2). 168–183.

Kiese-Himmel, Christiane. 2005. AWST-R: Aktiver Wortschatztest für 3- bis 5-jährige Kinder. Revision. Göttingen: Hogrefe-Testzentrale.

Kutas, Marta & Kara D. Federmeier. 2011. Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). Annual Review in Psychology 62. 621–647.

Laaha, Sabine, Dorit Ravid, Katharina Korecky-Kröll, Gregor Laaha & Wolfgang Dressler. 2006. Early noun plurals in German: regularity, productivity or default? Journal of Child Language 33. 271–302.

Leonard, Laurence B. 2009. Cross-linguistic studies of child language disorders. In Richard G. Schwartz (ed.), Handbook of Child Language Disorders, 308–324. Hove: Psychology Press.

Leonard, Laurence B., Patricia Deevy, Carol A. Miller, Leila Rauf, Monique Charest & Robert Kurtz. 2003. Surface forms and grammatical functions: Past tense and passive participle use by children with specific language impairment. Journal of Speech, Language, and Hearing Research 46(1). 43–55.

Levelt, Willem J. M., Ardi Roelofs & Antje S. Meyer. 1999. A theory of lexical access in speech production. Behavioral and Brain Sciences 22. 1–75.

Lindner, Katrin. 1998. Overgeneralization revisited: The case of German past participles. In Ray Fabri, Albert Ortmann & Teresa Parodi (eds.), Models of Inflection, 152–174. Tübingen: Niemeyer.

McCarthy, John. 2006. Prosodic morphology. In Keith Brown (ed.), Encyclopedia of Language and Linguistics, 182–190. Oxford: Elsevier.

McCarthy, John & Alan Prince. 1994. Prosodic morphology. In John Goldsmith (ed.), A Handbook of Phonological Theory, 318–366. Oxford: Oxford University Press.

Miller, Carol A., Laurence B. Leonard & Denise Finneran. 2008. Grammaticality judgements in adolescents with and without language impairment. International journal of language and communication disorders 43(3). 346–360.

Nespor, Marina & Irene Vogel. 2007. Prosodic Hierarchy. Berlin: De Gruyter. [First edition 1986 Dordrecht: Foris.]

Poll, Gerard H., Stacy K. Betz & Carol A. Miller. 2010. Identification of clinical markers of specific language impairment in adults. Journal of Speech, Language, and Hearing Research 53(2). 414–429.

Redmond, Sean M. 2003. Children's productions of the affix -ed in past tense and past participle contexts. Journal of Speech, Language, and Hearing Research 46(5). 1095–1109.

Rescorla, Leslie. 2002. Language and reading outcomes to age 9 in late-talking toddlers. Journal of Speech, Language, and Hearing Research 45(2). 360–371.

Rescorla, Leslie. 2005. Age 13 language and reading outcomes in late-talking toddlers. Journal of Speech, Language, and Hearing Research 48(1). 1–14.

Rescorla, Leslie. 2009. Age 17 language and reading outcomes in late-talking toddlers: support for a dimensional perspective on language delay. Journal of Speech, Language, and Hearing Research 52(1). 16–30.

Sabisch, Beate, Anja Hahne, Elisabeth Glass, Waldemar von Suchodoletz & Angela D. Friederici. 2009. Children with specific language impairment: The role of prosodic processes in explaining difficulties in processing syntactic information. Brain Research 1261. 37–44.

Schmidt-Kassow, Maren & Sonja A. Kotz. 2009. Event-related brain potentials suggest a late interaction of meter and syntax in the P600. Journal of Cognitive Neuroscience 21(9). 1693–1708.

Schwartz, Richard G. 2009. Specific language impairment. In Richard G. Schwartz (ed.), Handbook of Child Language Disorders, 3–43. Hove: Psychology Press.

Shahin, Antoine, Larry E. Roberts, Christo Pantev, Laurel J. Trainor &Bernhard Ross. 2005. Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. NeuroReport 16. 1781–1785.

Szagun, Gisela. 2001. Learning different regularities: The acquisition of noun plurals by German-speaking children. First Language 21. 111–141.

Szagun, Gisela. 2011. Regular/irregular is not the whole story: The role of frequency and generalization in the acquisition of German past participle inflection. Journal of Child Language 38(4). 731–762.

Wegener, Heide. 2002. Aufbau von markierten Pluralklassen im Deutschen – eine Herausforderung für die Markiertheitstheorie. Folia Linguistica 36. 261–295.

Weyerts, Helga, Martina Penke, Ulrike Dohrn, Harald Clahsen & Thomas F. Münte. 1997. Brain potentials indicate differences between regular and irregular German plurals. NeuroReport 8. 957–962.

Whitehouse, Andrew J. O., E. A. Line, Helen J. Watt &Dorothy V. M. Bishop. 2009. Qualitative aspects of developmental language impairment related to language and literacy outcome

in adulthood. International journal of language and communication disorders 44(4). 489–510.
Wiese, Richard. 2000. The phonology of German. Oxford, UK: Oxford University Press.
Wittek, Angela &Michael Tomasello. 2002. German children's productivity with tense morphology: The Perfekt (present perfect). Journal of Child Language 29(3). 567–589.

Gerrit Kentner

Schwa optionality and the prosodic shape of words and phrases

Abstract: This paper discusses effects of supra-lexical linguistic rhythm on the appearance or absence of optional schwa. Specifically, the roles of rhythmic alternation and prosodic parallelism are studied in three experiments and weighed against each other. In Experiment One, an oral reading study, readers were confronted with either of the two graphemic representations of the alternating adverb <gern(e)> ('happily') in sentential contexts the rhythmic structure of which was systematically varied. The evaluation of the scripted speech productions suggests that readers take the rhythmic environment into account when choosing an allomorph for the prosodically variable target word. Experiment Two is concerned with prosodic determinants for the morphosyntactic alternation in German partitive or possessive constructions. These may be realised as genitive attributes or using a prepositional construction. A forced choice experiment with written material suggests that participants consider the distribution of strong and weak syllables when choosing among the morphosyntactic variants. Experiment Three exploits the prosodic alternation of four adverbs. Analysing the distribution of the variants in a large written corpus attests that the immediate prosodic context affects the choice among the variants. A synopsis of the findings suggests that rhythmic alternation (conceived as the joint effects of stress clash avoidance and stress lapse avoidance) has a stronger impact on the presence or absence of a reduced syllable compared to prosodic parallelism.

Keywords: word prosody, optional schwa, prosodic morphology, German

---

Gerrit Kentner: Johann Wolfgang Goethe-Universität, Norbert-Wollheim-Platz 1, D-60629 Frankfurt am Main, kentner@lingua.uni-frankfurt.de

DOI 10.1515/9783110542899-006

1 Introduction

Several words in German are prosodically variable in that they may legally appear either with or without a schwa syllable – with no semantic effect associated with the presence or absence of the syllable headed by the reduced vowel. Schwa optionality is certainly a marginal phenomenon in German morphophonology, most likely because the appearance of schwa is usually morphologically governed, with schwa corresponding to the exponent of e.g. a plural, first person, or agreement morpheme. In spite of schwa's role as inflectional exponent, schwa optionality is attested in all major word classes, as the list in (1) reveals.

(1) a. Nouns: *die Tür ~ die Türe* ('the door')
  b. Verbs (first person sg., pres.): *ich geh ~ ich gehe* ('I go')
  c. Predicative adjectives: *fad ~ fade* ('dull, tasteless')
  d. Adverbs: *gern ~ gerne* ('happily')
  e. Demonstrative pronouns: *dies ~ dieses* ('this')
  f. Conjunctions: *eh' ~ ehe* ('before')
  g. Prepositions: *ohn' ~ ohne* ('without')
  h. Numerals: *zu zweit ~ zu zweien* ('two by two', 'in sets of two')

A variety of factors are known to impinge on the appearance or absence of optional schwa syllables, among them language change, speaking rate and style (or register), and dialectal influence. That is, not all alternating forms in (1) are equally likely to occur in modern Standard German. For instance, the dated numeral (1h.) of the form *zu NUM-en* has by now been almost fully replaced by the current schwa-less construction *zu NUM-t*. Monosyllabic *ohn'* (1g.) is confined to certain poetic registers, while (1c.) has a dialectal distribution. The appearance of optional schwa (or its orthographic cognate <e>) in genitives *Jahrs ~ Jahres* ('year') has been shown to be more likely the higher the frequency of the noun is (Fehringer 2011). Still, some alternating forms seem to happily coexist and vary almost freely even within the same historical and dialectal strata.

Aside from factors like usage frequency, speech register, and dialectal distribution, the prosodic-phonological context the variable word is embedded in has been discussed as potentially conditioning the distribution of forms with or without schwa syllable. Studies by Rohdenburg (2014), Schlüter (2005), and Wiese and Speyer (2015) suggest that the prosodic makeup of adjacent words may co-determine the choice among the prosodically varying allomorphs. The claim put forward in these studies is that speakers exploit schwa-optionality to improve the phrasal rhythm. There are at least two, partly conflicting, ways in which phrasal rhythm may be improved. On the one hand, speakers may, whenever possible, strive for an alternation of stressed (or strong) and unstressed (or weak) syllables, thereby creating a beat that is as regular as possible. This entails that structures involving sequences of adjacent stressed syllables (stress clash) or sequences of unstressed syllables (stress lapses) are disfavoured. On the other hand, the rhythmicity of an utterance may be enhanced by iterating prosodic units of the same type, fostering prosodic parallelism. Accordingly, a prosodic phrase that consists of two words is favoured if the two words exhibit the same prosodic structures (e.g. either two monosyllabic words or two trochees); a sequence of two prosodically different words (e.g. a trochee followed by a monosyllable) would violate the iterative rhythm.

This paper takes a fresh look at the various effects of supra-lexical linguistic rhythm on the appearance or absence of optional schwa. Specifically, the roles of rhythmic alternation on the one hand, and iterative rhythm or prosodic parallelism on the other will be studied in three experiments and weighed against each other. In Experiment One (Section 2.1), an oral reading study, readers were confronted with either of the two graphemic representations of the alternating adverb <gern(e)> ('happily') in sentential contexts that were systematically varied with respect to rhythmic structure. The evaluation of the scripted speech productions suggests that readers take the rhythmic environment into account when reading out the written target word. Experiment Two (Section 2.2) is concerned with prosodic determinants for the morphosyntactic alternation in German partitive or possessive constructions. These may be realised as genitive attributes or prepositional constructions. A forced choice experiment with written material suggests that participants consider the distribution of strong and weak syllables in the possessum when choosing among the morphosyntactic variants, confirming a rhythmic-prosodic effect. Finally, Experiment Three (Section 2.3) exploits the prosodic alternation of the adverbs *gern* ~ *gerne*, *lang* ~ *lange*, *selbst* ~ *selber*, *meist* ~ *meistens* ('happily, for a long time, self, most of the time'). Analysing the distribution of the variants in a large written corpus attests that the propensity for rhythmic alternation affects the choice among these variants. A synopsis of the endings suggests that rhythmic alternation (conceived as the joint effects of stress clash and stress lapse avoidance) has a stronger effect on the presence or absence of a reduced syllable compared to prosodic parallelism. Before reporting on the experiments in Section 2, the remainder of Section 1 provides relevant background on prosodic structure and linguistic rhythm in German (and beyond).

1.1 (Supra-)lexical prosodic structure and linguistic rhythm

As for word-internal prosody, the core of the German lexicon and morphological system is prosodically constrained in that it displays a strong preference for disyllabic, trochaic forms (for a review, see Domahs, Domahs, and Kauschke 2018, this volume). The trochaic preference dictates e.g. the choice of plural allomorphs (Eisenberg 1991; Wegener 2004; Wiese 2009), and it restricts the productivity of many derivations, such as umlaut in diminutive formation (Fanselow and Féry 2002) or the possibility to form denominal adjectives by suffixation of *-ig* (this derivation is only licit when the suffix is immediately preceded by a syllable carrying stress, thus forming a right-aligned trochee[1]: *ruhig* < *Ruhe*, *tomatig* < *Tomate*, **kürbisig* < *Kürbis*, **mangoig* < *Mango*, **paprikaig* < *Paprika*[2]). The effect of the trochee in German morphology is probably best seen in hypocoristic truncations with the *i*-suffix (*Ándi* < *Andréas*, *Stúdi* < *Studént*; cf. Féry 1997; Itô and Mester 1997; Köpcke 2002) in which the trochaic template applies almost exceptionless – in fact, as the examples *Andi* and *Studi* show, this highly productive process may even force the deviance from the stress pattern of the source form to safeguard a trochee. The trochee may thus be understood as an optimal template regulating the shape of words.

Beyond the word, the trochee may lead to rhythmic alternation of strong and weak syllables. In the ideal case, the concatenation of words yields a concatenation of trochees and, consequentially, the perfect alternation of strong and weak beats. A trochaic structure like (2) fulfills pertinent conditions regarding rhythmic alternation, namely the constraints against clustering of strong syllables (*Clash; see Anttila et al. 2010, for various instantiations of this constraint) or against sequences of weak syllables (*Lapse; cf. Shih et al. 2015, for a discussion of different eurhythmy measures). The example in (2) can be considered especially eurhythmical in that the alternation between strong and weak is even reflected in the vowel qualities with diphthongs or long vowels alternating with unstressable reduced vowels.

1 In the case of stems ending in a syllabic sonorant, a schwa syllable may be skipped, producing dactylic forms like *hibbelig* ('jittery'). This is reminiscent of Kager's notion of the invisibility of schwa syllables to certain phonological processes that are conditioned by stress (Kager 1989), as may be exemplified by German umlaut (Féry 1994).

2 A reviewer suggests that hiatus avoidance may be considered a factor in the ungrammaticality of **mangoig* and **paprikaig*. However, cases like *ruhig* [ʁuːɪç] and *böig* [bøːɪç] < *Böe* ('gust', 'squall') attest the license for hiatus in these contexts.

(2) *Friede, Freude, Eierkuchen*
[(ˈfʁiː.də) (ˈfʁɔɪ.də) (ˈʔaɪ.ɐ) (ˌkuː.xən)]
peace, joy, pancake
'love, peace and harmony'

Apart from the alternating rhythm of strong and weak syllables, (2) constitutes a prime example for iterating rhythm (or prosodic parallelism) with the four trochees building a perfectly parallelistic prosodic structure, i.e. a symmetric set of two pairs of trochees. The prosodic repetitiveness is enhanced by the segmental structure at least for the first pair of trochees (the parallelism is mirrored in the onsets of both the stressed and the unstressed syllable). The alliterating idiomatic expression in (2) suggests that the force of iterating rhythm is most obvious in poetic language where prosodic parallelism is prevalent (concerning, for example, the matching of lines in metered poems, see Menninghaus et al. 2017).

1.2 Rhythmic alternation within and beyond the word

The propensity for rhythmic alternation (i.e. the effect of *CLASH and *LAPSE) is illustrated by cases in which it forces a deviance from patterns that would be expected by mere concatenation of morphs. For instance, the prominence of syllables can be demoted to avoid a clash of neighbouring strong beats. Consider, in this respect, the German word *Nation* [naˈtsjoːn] ('nation') with the latinate suffix *-ion* attracting stress on the final syllable. Attaching the equally stress-attracting adjectival suffix *-al* leads to a restructuring of prominences such that the stem-final syllable becomes unstressed and instead the initial syllable receives secondary stress (*national* [ˌnatsjoˈnaːl]).

In other cases, the force of *CLASH may even impinge on the quality of the underlying vowel. This is the case in the most natural rendition of a compound like *Bauarbeiter* 'builder', made up of the constituents *Bau* [baʊ] 'building' and *Arbeiter* [ˈʔaː.baɪ.tɐ] 'worker'. With compound stress on the first constituent, the initial syllable of the second member becomes a reduced syllable and is thus attached to the foot projected by the monosyllabic first member [(ˈbaʊɐ)(ˌbaɪ.tɐ)]. Importantly, the footing of this compound, arguably driven by *CLASH, runs counter to its morphological structure.[3]

---

3 It is certainly possible for *Bauarbeiter* to retain secondary stress on the first syllable of the head noun and, in addition, mark the morphological boundary by a glottal stop [(ˈbaʊ)(ˌʔaː.baɪ.tɐ)]. I would argue, however, that this rendition is only valid under a strong

The avoidance of clashes has also been shown to have syntactic effects (cf. especially Schlüter 2005; Speyer 2010, on the syntactic effects of clash avoidance in English). As for German, consider the otherwise unmotivated ordering with the adverbial intensifier *ganz* or *so* that is separated from the adjective or de-adjectival noun it modifies and instead preceding the indefinite pronoun or determiner in noun phrases like (3) (see e.g. Kallulli and Rothmayr 2008 and Gutzmann and Turgay 2015 for syntactic and semantic analyses of similar phenomena). This inversion coexists alongside the canonical ordering with the determiner preceding the intensifying adverb. As noted by Behaghel (1930), the displaced determiner serves as a buffer between two prominent syllables, preventing a clash. Determiner doubling in (4) provides an even more striking case, arguably with the same motivation.[4]

(3) canonical order ~ determiner inversion
  a. *was ganz Neues ~ ganz was Neues*
     'something quite new' ~ 'quite something new'
  b. *ein ganz junger Mann ~ ganz ein junger Mann*
     'a quite young man' ~ 'quite a young man'

(4) determiner doubling
  a. *ein ganz ein feiner Kerl*
     'a quite a fine chap'
  b. *ein so ein großer Bub*
     'a such a big boy'

Interestingly, inversion or doubling appears to be illicit in German with di- or trisyllabic intensifiers (*gänzlich*, *dermaßen*) whose unstressed final syllable prevents a stress clash in the first place.

(5) a. *ein gänzlich feiner Kerl*
  a.' **gänzlich ein feiner Kerl*
  a." **ein gänzlich ein feiner Kerl*
     '(a) quite (a) fine chap'

---

pragmatic pressure to clarify the morphological structure (e.g. in the case of a misunderstanding or correction), and uncommon in running speech.

4 Schlüter (2005) notes the same inversion of the indefinite determiner and the adverb *quite* in English and argues that this inversion has a rhythmic motivation.

b. *ein dermaßen großer Bub*
b.' **dermaßen ein großer Bub*
b." **ein dermaßen ein großer Bub*
'(a) such (a) big boy'

The ungrammaticality of inversion or doubling in (5) casts doubt on purely syntactic accounts of this phenomenon and instead provides further evidence for a rhythmic trigger for these word order options.

As noted above, apart from stress clashes, sequences of unstressed syllables are considered disrhythmic and are thus avoided.[5] For instance, when the productive suffix *-er* is attached to trochaic place names ending in *-en* [ən] to derive a demonym to the place name, elision of a reduced syllable is common in certain dialects.

The elision of a reduced syllable in these dialects is probably motivated by linguistic rhythm, specifically to avoid sequences of two reduced syllables (*LAPSE). This process seems to have an areal distribution such that it does not affect all place names in the same way, as may be observed when comparing (6) and (7) with (8).[6]

(6) semi-transparent, with resyllabification of stem-final consonant(s) (mainly East Central German and Bavarian)
a. *Dresden* [dʁeːs.dən] – *Dresdner* [dʁeːs.dnɐ]
b. *Bautzen* [baʊ.tsən] – *Bautzner* [baʊ.tsnɐ]
c. *München* [mʏn.çən] – *Münchner* [mʏn.çnɐ]
d. *Weiden* [vaɪ.dən] – *Weidner* [vaɪ.dnɐ]

(7) opaque, elision of stem-final consonant (Northern Low Saxon)
a. *Emden* [ʔɛm.dən] – *Emder* [ʔɛm.dɐ]
b. *Bremen* [bʁeː.mən] – *Bremer* [bʁeː.mɐ]
c. *Norden* [nɔɐ.dən] – *Norder* [nɔɐ.dɐ]
d. *Apen* [ʔaː.pən] – *Aper* [ʔaː.pɐ]

5 The Strict Layer Hypothesis assumed in prosodic phonology (Selkirk 1984) provides a supplementary explanation for the avoidance of lapses: under the assumption that feet in German are trochaic and maximally disyllabic, further unstressed syllables cannot be parsed into feet and thus constitute a violation of the principle EXHAUSTIVITY.
6 The examples represent the written norm within the dialectal areas. For certain place names, reduced and full form coexist, e.g. *Uelzen – Uelzener ~ Uelzer*; *Dülmen – Dülmener ~ Dülmer*.

(8) fully transparent, no effect of *LAPSE
  a. *Hagen* [haː.gən] – *Hagener* [haː.gə.nɐ]
  b. *Siegen* [ziː.gən] – *Siegener* [ziː.gə.nɐ]
  c. *Gießen* [giː.sən] – *Gießener* [giː.sə.nɐ]
  d. *Aachen* [ʔaː.xən] – *Aachener* [ʔaː.xə.nɐ]

The effect of *LAPSE is especially remarkable when considering dactylic place names like *Tübingen*, *Kaufungen*, *Bevensen*. Mere affixation of the demonymic suffix would lead to three consecutive unstressed syllables, a configuration that is ungrammatical across dialects (**Tübingener*, **Kaufungener*, **Bevensener*). Instead, *LAPSE dictates haplology in these cases, resulting in elision of stem-final [ən] – as in (7) – to yield *Tübinger*, *Kaufunger*, *Bevenser*.

Vogel et al. (2015) have shown clear effects of *LAPSE on the linearization of constituents in German sentences. Specifically, Vogel et al. (2015) investigated i) the ordering of inherently weak pronominal adverbs in the Middlefield and ii) auxiliary verbs in sentence final verb clusters. Speakers were to repeat sentences with these constructions that were presented in either a rhythmically alternating or a disrhythmic condition, i.e. one in which the placement of the weak pronoun or auxiliary verb leads to three consecutive, unstressed syllables. The results reveal a clear effect of rhythm such that recall errors were significantly more likely in the disrhythmic conditions. In other words, the syntactic representation of the sentences to be recalled was more stable when the corresponding prosodic representation was rhythmically optimal.

The preceding examples attest the importance of rhythmic alternation, more specifically, of the constraints *CLASH and *LAPSE for the phonological representation (and processing) not only of words but also at the phrasal level. The low level constraints fostering an alternating rhythm of strong and weak syllables are complemented by a tendency for iterating rhythm such that patterns that emerge from the distribution of prominences are preferably repeated.

1.3 Prosodic parallelism within and beyond the word

Recently, Wiese and Speyer (2015) suggested that prosodic parallelism is relevant for the occurrence of final schwa in cases like (1) (see Kentner 2015, for discussion). In a nutshell, the idea is as follows: when given the choice, speakers strive for prosodic parallelism; for two words that are prosodic phrase mates, the foot structures are preferably parallel, i.e. the feet display the same number of syllables and stress pattern. Thus, their argument goes, the appearance or lack of optional schwa is dependent on the foot structure of neighboring words.

Examining a large corpus of written German, Wiese and Speyer (2015) investigated, inter alia, several cases of nouns with apparently freely alternating monosyllabic and disyllabic variants like *Tür* – *Türe* ('door') or *Tags* – *Tages* ('dayGEN') in the context of (preceding) monosyllabic or disyllabic determiners.

(9) a. $((\text{die})_\Sigma\ (\text{Tür})_\Sigma)_\varphi$
b. $((\text{ei.ne})_\Sigma\ (\text{Tü.re})_\Sigma)_\varphi$

Using chi-square tests on bigram frequencies, they disprove statistical independence of the prosodic shapes of co-occurring determiner and noun. In a follow-up study, Wiese (2016) reports corroborating evidence in corpora of spontaneous spoken language. These results suggest that, when possible, the prosodic structure of the noun preferably mirrors the structure of the determiner, cf. (9). Note that this explanation assumes that function words like determiners project a foot (see Kentner 2015, for discussion).

The effect of prosodic parallelism is not confined to German schwa-zero alternations alone. In fact, there are phenomena that would defy proper analysis without recourse to a constraint on prosodic parallelism; these are cases in which the PARALLELISM constraint appears to have a stronger influence compared to the German schwa-zero alternations, in which parallelism is merely a tendency. Consider Standard Chinese, in which the productivity of N+N compounds and V+Obj combinations is strictly constrained by the number of syllables. As Duanmu (2012) shows, parallel prosodic structures with either two monosyllables (1+1) or two disyllables (2+2) are generally licit for both constructions. However, for N+N compounds, non-parallel structures of the 1+2 type are mostly unacceptable. Similarly, for V+Obj phrases, the imbalanced pattern 2+1 is considered unacceptable (cf. Luo and Zhou 2010; Luo, Duan, and Zhou 2015, for pertinent neuro- and psycholinguistic evidence).

Another case demonstrating the influence of PARALLELISM, again in German morphophonology, is rhyme and ablaut reduplication (Kentner 2017). This type of reduplication has a strict non-identity requirement concerning base and reduplicant, both of which correspond to a prosodic foot (*schickimicki*, **schickischicki* < *schick* 'posh'; *hickhack*, **hackhack* < *hacken* 'to chop, to bicker'). Crucially, nonidentity is confined to the segmental tier. That is, a difference between base and reduplicant concerning the prosodic shape is prohibited (**schischicki*, **schickischick*; **hickhacke*, **hickehack*), and it is this prohibition that strongly suggests the workings of prosodic parallelism.

Wiese and Speyer's proposal on prosodic parallelism is in line with the observation that equal-sized prosodic constituents are preferred on various levels of prosodic representation and processing. This finding has been codified in several ways: for instance, Ghini (1993) suggests that, in Italian, prosodic structure is built in a fashion that guarantees balanced phonological phrases even if the resulting phrasing is non-isomorphic to syntactic structure. Similarly, Myrberg (2013), examining Stockholm Swedish, suggests that prosodic subconstituents conjoined within an intonational phrase preferably have the same prosodic status. Selkirk (2000) invokes the constraints BinMin and BinMax which jointly favor minimally and maximally two prosodic words per phrase. Féry and Kentner (2010) and Kentner and Féry (2013) propose a Similarity condition on prosodic structure such that neighboring constituents at the same level of syntactic embedding be adjusted to exhibit a similar prosodic rendering, irrespective of the constituents' inherent complexity.

Given the abundance of evidence for parallelism, it is not far-fetched to consider it a well-formedness condition on prosodic structure, just like *Clash and *Lapse. The exact formulation of this constraint, however, is open to debate (as is the formulation of *Clash and *Lapse, cf. discussions in Anttila, Adams, and Speriosu 2010; Shih et al. 2015). Suffice it to say that the Parallel constraint requires adjacent prosodic constituents grouped within a higher constituent to exhibit the same prosodic structure.

Having introduced the three rhythmic constraints and their workings in various environments, the following section assesses their relative contribution to word prosodic structure and phrasal rhythm.

2 Three studies on word prosodic structure and phrasal rhythm

The three studies to be presented below were designed to explore the influence of the rhythmic environment on morphophonological (and morphosyntactic) variation in German. Although prosody (or particularly prosodic rhythm) is not explicitly encoded in the written modality (but see Evertz and Primus 2013), all three experiments use written material for this purpose. This is justified by numerous findings which converge to suggest that the use of the written modality (reading and writing) involves recourse to prosodic representations (see, e.g. Chafe 1988; Breen 2014; and the collection of studies in Kentner and Steinhauer 2017).

The experiments use different linguistic environments and employ different methodologies but all share as crucial factor the distribution of lexically strong or stressed syllables around the morphophonologically or morphosyntactically variable word(s). Experiment One is an oral reading experiment that focuses on the prosodic rendering of the variable prosodic adverb *gern(e)* in different rhythmic contexts. A large-scale online survey (>150 participants), Experiment Two explores rhythmic influences on the choice between possible realisations of possessive or partitive relations. In Experiment Three, we return to prosodically variable adverbs. Employing a corpus analysis, we investigate the usage frequency of prosodically variable adverb-verb sequences to specifically pit effects of rhythmic alternation (avoidance of stress clash and lapse) against those of prosodic parallelism.

2.1 Rhythmic context effects on optional schwa in read speech

The first experiment is concerned with the effects of the rhythmic-prosodic context on the realisation of the prosodically variable adverb *gern(e)* ('happily') in spontaneous (unprepared) oral reading. This adverb has two graphemic representations that correspond to i) a monosyllabic <*gern*> or ii) a trochaic variant <*gerne*>. For the experiment, both graphemic variants were embedded in sentences with systematically varied rhythmic-prosodic structures to ascertain the effect of the rhythmic context on the realisation of schwa on the adverb in scripted speech production.

Previous work suggests that optional schwa syllables are used by speakers to optimise the rhythmicity of phrases and sentences; specifically, it has been argued that a schwa syllable may act as a buffer syllable that prohibits stress clash (Kuijpers and van Donselaar 1998; Rohdenburg 2014; Schlüter 2005). In the case of the variable adverb <*gern(e)*>, the optional schwa syllable may thus prevent a potential clash with a word to the right of it.

The first manipulation of this experiment therefore targets the syllable to the right of the variable word: the noun following the variable adverb in (10) begins in either a stressed (*Hímbeeren*) or an unstressed syllable (*Kartóffeln*). In addition, the rhythmic context to the left of the word was manipulated; this manipulation is motivated by the hypothesised propensity for iterating or sequential rhythm that is at the core of prosodic parallelism. The lexical material of the sentences was constructed to yield a trochaic beat with every other syllable bearing lexical stress. The syllabic structure of the noun directly preceding the target adverb <*gern(e)*> was systematically varied, with either a monosyllable (*Hof*) or a disyllabic trochee (*Garten*). Thus, the first (or only) syllable of the variable adverb falls on either an on-beat or off-beat position of the established trochaic pattern. According to the parallelism hypothesis, the trochaic form of the variable adverb should be preferred when preceded by a trochee while the monosyllabic form should be preferred when preceded by a monosyllabic foot.

(10) a. *Bodo will in Steffis Garten gerne Himbeeren ernten.*
b. *Bodo will in Steffis Garten gern Himbeeren ernten.*
c. *Bodo will in Steffis Hof gerne Himbeeren ernten.*
d. *Bodo will in Steffis Hof gern Himbeeren ernten.*
e. *Bodo will in Steffis Garten gerne Kartoffeln ernten.*
f. *Bodo will in Steffis Garten gern Kartoffeln ernten.*
g. *Bodo will in Steffis Hof gerne Kartoffeln ernten.*
h. *Bodo will in Steffis Hof gern Kartoffeln ernten.*
'Bodo would like to harvest {raspberries, potatoes} in Steffi's {yard, garden}'

2.1.1 Materials, participants, procedure

Twenty-four item sets like (10) were devised. The items were distributed over eight lists such that items and conditions were counterbalanced across the lists with each list containing exactly one condition from each item set. Additionally, each list contained 64 filler items from four unrelated experiments and three practice items not connected to any of the experimental items, yielding a total of 91 items. With the exception of the three initial practice items, the item order was determined by pseudo-randomization (van Casteren and Davis 2006) for each participant individually such that items from the same experiment had a minimal distance of two intervening items from other experiments and items from the same experimental condition were separated by at least three fillers.

Twenty-four members (19 female) of the Goethe-University community (Frankfurt, Germany) took part in the experiment. All participants are native speakers of German with normal or corrected-to-normal vision per self report. Initially, participants were not informed about the purpose of the experiment but debriefed after the experiment ended.

The experiment took place in a silent office at Goethe University in single sessions for each participant. Participants were seated in front of a 21.5-inch computer screen and equipped with a microphone head set (Shure) attached to an R-44 digital recorder.

All 91 items of each list were presented in a slide show. Each item was presented on two consecutive screen displays. The first display presented two (irrelevant) context sentences in the upper half and the first two words of the target sentence (in the case of this experiment: subject and modal verb) in the middle of the screen (all text left-aligned). Upon pressing the enter button on the keyboard, the target sentence appeared in full (leaving the rest of the first display intact). Participants were asked to read the first display (i.e. the context) silently before moving on to the second display screen. To ensure spontaneous, unprepared oral reading and minimal look-ahead, participants were instructed to read out the target sentence immediately as it appeared on screen and to do so as fluently as possible. The participants were discouraged from making corrections during or after reading and to move on to the next item after reading by another button press. The productions of the participants were recorded on a digital memory card.

2.1.2 Results

All in all, (24 items x 24 participants =) 576 experimental sentences were recorded. Two student assistants independently evaluated each target sentence. Their task was to determine by ear i) whether the production was a fluent and flawless response to the target sentence and ii) whether the target adverb was realised as monosyllablic *gern* or disyllabic *gerne*.

Seven sentences (1.2%) were scored as non-fluent or otherwise flawed by at least one referee and discarded from further analysis. The judgments concerning the number of syllables were perfectly consistent. Aggregating the 569 valid responses, the adverb was judged to contain a reduced syllable in 260 cases (45.7%) and monosyllabic in 309 cases (54.3%) suggesting a slight preference for the monosyllabic form. All in all, the oral realisation of the adverb corresponded to the graphemic representation in 82% of the cases.

Mixed logistic models (Bates et al. 2013) were applied in the statistical computing environment R (R Core Team 2015) to assess the effects of the graphemic representation ('writtenE'), the rhythmic environment to the left ('RhythmLeft'), and the rhythmic environment to the right ('RhythmRight') on the realisation of the schwa syllable (dependent variable: 'realiseE') in reading. The fixed effects (or predictor variables) were coded as orthogonal sum contrasts to ensure minimal correlation. Apart from the fixed effects, the models included participant ('speaker') and item as random effects that were adjusted for by participant or by item differences in the effects of the predictor variables. Likelihood ratio tests (carried out by the anova function) were used to compare models with different predictor variables and random effect structures in order to determine the model with the best fit for the data. The likelihood ratio test generally prefers simpler or more parsimonious models over more complex ones as long as the inclusion of model parameters does not significantly increase model fit. Consequentially, if the inclusion of a model parameter did not significantly improve model fit, it was culled from the model. Complex models with all three predictor variables, the respective interactions, and complex random effects structures[7] were tested first and non-significant predictors (as determined by the likelihood ratio tests) were culled in a stepwise fashion. Over and above a highly significant effect of the graphemic representation (readers preferably realise the adverb in line with its graphemic representation), the preceding context significantly affects the realisation of the reduced syllable. As visible in Table 1, trochaic *gerne* appears to be more likely when the preceding word is trochaic; conversely, the monosyllabic variant is preferred after monosyllabic nouns. The coefficients of the best fitting logistic mixed model are tabulated in Table 2.

Tab. 1: Percentages of trochaic realisations of the variable adverb *gern(e)* broken down by graphemic form of stimulus and prosodic form of preceding noun

Percentage of trochaic realisations of adverb		**Prosodic form of noun preceding the adverb**	
		trochaic	**monosyllabic**
Graphemic form of the adverb	<gern>	17	10
	<gerne>	80	76

Tab. 2: Coefficients of the best fitting mixed logistic model with the formula *glmer(realiseE~writtenE+RhythmL+(writtenE | speaker)*, family = binomial). N = 569

	Estimate	**Std. Error**	**z-value**	**p-value**
(Intercept)	−0.06767	0.43364	−0.156	0.8760
writtenE	2.50456	0.33543	7.467	<0.001
RhythmLeft	−0.27643	0.13822	−2.00	0.0455

7 Several of the more complex models did not converge. Non-converging models were not considered further in the model comparison process.

2.1.3 Discussion

The experiment shows that readers are generally guided by the written form of the prosodically variable word when producing it in spontaneous read speech. Apart from the effect of the graphemic representation, the rhythmic context has a small but significant effect on the realisation of schwa on the critical adverb. This, however, only holds for the manipulation concerning the rhythmic structure to the left of the critical word (RhythmLeft). The other rhythmic effect that was tested in this experiment, the rhythmic context to the right of the critical word, failed to affect the realisation of *gern(e)*. One conceivable explanation is related to the task of spontaneous oral reading: readers may simply not have had the time to sufficiently process the upcoming word to prosodically adjust the target word to it.

The significant effect of RhythmLeft suggests that readers prefer monosyllabic *gern* after a monosyllabic noun while trochaic *gerne* preferentially follows a trochaic noun. This finding, at first sight, corroborates the prediction according to the parallelism hypothesis. However, taking into account the wider prosodic context (with the trochaic beat that was established right from the beginning of the sentence), parallelism as formulated by Wiese and Speyer (2015) may be insufficient to explain the results. Under Wiese and Speyer's account, and under the Strict Layer Hypothesis (SLH) of prosodic phonology (Selkirk 1984), feet cannot straddle word boundaries. This limitation, however, is crucial when evaluating the parallelism effect. Compare, in this respect, the conflicting footings of an example item in (11):

(11) Conceivable foot structures
  a. trochaic footing ('Abercrombian' feet)
    (Rosie) (will auf) (jeden) ***(Fall gern)*** *(Ärztin) (werden)*
  b. footing according to the Strict Layer Hypothesis
    (Rosie) (will) (auf) (jeden) ***(Fall) (gern)*** *(Ärztin) (werden)*
    Rosie wants in any case happily physician become
    'In any case, Rosie would like to become a physician.'

(11a.) represents a perfectly iterating prosodic structure – a sequence of six trochees – but blatantly violates the Strict Layer Hypothesis, with the adverb *gern* demoted to the weak position of a trochee (in this position presumably being unaccentable); (11b.), in contrast, abides by the letter of the SLH but the structure fails to represent the trochaic beat that is felt when the sentence is uttered. This is because, according to Wiese and Speyer (2015), even function words project feet (see Kentner 2015, for criticism). If one were to follow Wiese and Speyer (2015), the intended trochaic beat of the experimental items does not correspond to parallel prosodic structures in the first place.

The analysis of the read sentences so far only considered the presence or absence of schwa on the critical adverb but did not involve any assessment of its prosodic prominence. A cursory look at the realisations of the adverb, however, suggests that the monosyllabic adverb often remains entirely unaccented (which would be in line with the representation in (11a.)) and often features a centralised vowel: [gən]. There is independent evidence to the effect that leaving the adverb unaccented (a necessity for the representation in (11a.)) is very common: Kutscher (2016) found that adverbs in German are often prosodically reduced, and thus serve as a trough between prominence peaks, preventing stress clash.

While I acknowledge that this experiment cannot settle the largely theoretical debate among the schools favoring Abercrombian feet (11a.) over those abiding by the SLH (11b.) or vice versa, I point out that the representation (11a.) not only respects *CLASH and *LAPSE; (11a.) also exhibits a sequential rhythm and may thus be in line with a weaker version of PARALLELISM that tolerates violations of the SLH. (11b.), in contrast, only locally fulfills the PARALLELISM constraint (in the bolded part of the sentence) but fails to respect other constraints on rhythmic structure (*CLASH, *LAPSE) in spite of the fact that a natural rendition of the sentence exhibits a perfect alternation of prominences.

2.2 Rhythm and morphosyntactic choice: Morphological genitive vs. prepositional construction

In German, the possessive or partitive relationship may be expressed by (at least) two syntactically distinct constructions:[8] by morphological case (genitive) or by a prepositional phrase headed by *von* ('of'). The choice between these two is partly governed by register or style with the prepositional construction deemed more colloquial and the genitive more formal. Given that the prepositional construction affords more (function) words than the morphological genitive, the two variants also exhibit a difference concerning their rhythmic patterns.

8 In the following, further options will be disregarded, e.g. compounding *Herbergswirt* (lit. 'inn owner') or the preposed genitive, as in *Marias Hund* ('Maria's dog'). The latter construction is confined to animate genitives and mainly used with proper names.

(12) a. *Der Wirt der Herberge*
the owner the.GEN inn.GEN
b. *Der Wirt von der Herberge*
the owner of the.DAT inn.DAT
'the owner of the inn'

As apparent from a comparison of the two syntactic options in (12), the prepositional phrase (12b.) involves, in addition to the determiner, a (usually) unaccented syllable (the preposition *von*) which increases the distance between head noun and attribute. When the latter two are lexical words and new to the discourse context, these referents usually bear an accent. The exact location of the two accents and their distance from each other depends not only on the construction (genitive or prepositional phrase) but also on the prosodic structures of the lexical words involved. The accents are the further apart the more unstressed syllables follow the head noun's stressed syllable, or the more unstressed syllables the attributive noun has preceding its stressed one.

Making use of a systematic manipulation of the prosodic distance between head noun accent and accent on the attribute, the following study aims at testing the hypothesis that the choice between the two syntactic options is attributable to the (implicit) rhythmic structure they engender. A very similar hypothesis has recently been confirmed for the usage of the English *s*-genitive and 'of'-genitive by Shih et al. (2015) who conducted a large-scale analysis using a corpus of spoken American English; in their dataset, however, the effects of rhythm on construction choice, although detectable, were largely dampened by the factor animacy. Here, a more controlled experimental avenue was chosen, i.e. a questionnaire study in which the prosodic structures of both the head noun and the attribute were systematically varied while leaving the factor animacy constant. The study will be detailed in the following.

2.2.1 Materials and method

An online questionnaire (Sosci Survey by Leiner 2014) was set up in which participants had to tick their preferred option for the expression of a possessive or partitive relationship in various rhythmic conditions. To this end, 24 items like (13a.–d.) were devised with head noun and attribute separated by a blank. The four conditions of the 24 items were counterbalanced across four blocks in a latin square design such that no head noun and attribute was presented more than once per block. The trials were presented in randomised order, interspersed with 40 filler items from two unrelated experiments. Each item was presented on a single slide together with four options to fill the blank. By ticking the appropriate box, participants had to choose either *der* (i.e. the monosyllabic definite determiner for the genitive attribute) or *einer* (i.e. the disyllabic indefinite determiner for the genitive attribute) or *von der* (i.e. the preposition and following determiner for the prepositional construction). In addition, a fourth option (*aus* 'from') was given as an oddball option that invariably leads to an ungrammatical construction. This was included to be able to spot participants who randomly marked one option without proper consideration of the item.

175 students of the Goethe-University community participated in the online questionnaire. Each participant was randomly assigned to one block.

(13) Insert *der* or *einer* or *von der* or *aus*
a. *Der Knopf ... Arbeitshose*[9]
b. *Die Knöpfe ... Arbeitshose*
c. *Der Knopf ... Gesäßtasche*
d. *Die Knöpfe ... Gesäßtasche*
'the button(s) {a., c.: Sg; b., d.: Pl} of the {a., b.: work pants; c., d.: back pocket}'

2.2.2 Predictions

The study was originally designed to test the hypothesis that construction choice is co-determined by the propensity for rhythmic alternation. Correspondingly, more short genitives (the monosyllabic determiner *der*) are expected in conditions with greater distance between the accented syllables of head noun and attribute, i.e. when the head noun has non-final stress and the attribute has non-initial stress. Conversely, more prepositional constructions *von der* or disyllabic genitives *einer* are expected in conditions with a short distance between the accented syllables of head noun and attribute. Opposing predictions come about when considering effects of prosodic parallelism. According to the PARALLELISM constraint, structures are preferred that yield an iterating rhythm. Correspondingly, in our case, a trochaic head noun (such as *Knöpfe*) should give rise to a preference for the disyllabic determiner *einer* or the prepositional construction with *von der* (the monosyllabic preposition and monosyllabic determiner

---

9 All attribute nouns have feminine gender in order to avoid fusion of preposition and determiner, a common process with masculine or neuter attributes (*von dem* > *vom*).

are assumed to be grouped into a trochee). A monosyllabic head noun, in turn, should promote the monosyllabic determiner der.[10]

Tab. 3: Percentages for chosen possessive/partitive construction broken down by stress on head noun and attribute

		Ultima of head noun		Initial of attribute noun	
		unstressed	**stressed**	**unstressed**	**stressed**
Prosodic form of Gen or PP	monosyllabic	59	55	58	56
	trochaic	41	45	42	44
	Total	100	100	100	100

2.2.3 Results

Several participants only partly completed the questionnaire, resulting in many missing answers. All in all, 3662 responses or 87% of the expected 4200 (= 175 participants x 24 items) were collected. In 39 of the cases, the oddball option *aus* was chosen, resulting in ungrammatical constructions. The majority (90%) of the remaining 3623 valid responses resulted in a genitive construction (2055 times or 57% of the cases monosyllabic *der*; 1195 times or 33% of the cases disyllabic *einer*). In only 373 or 10% of the cases, the prepositional construction was chosen. One reason for this discrepancy lies in the fact that there were two options to choose from genitives but only one valid prepositional option (not counting the ungrammatical oddball). Moreover, since the task was presented in writing, there is certainly a tendency to choose the formal genitive over the more colloquial prepositional construction.

In order to specifically test the predictions according to the principle of rhythmic alternation (*CLASH, *LAPSE) and PARALLELISM, the responses were grouped by prosodic structure, i.e. the disyllabic trochaic genitive determiner

10 The prosodic structure of the attribute was varied in such a way as to make predictions according to prosodic parallelism impossible to test with respect to the attribute noun. The first syllable of the attribute was either a stressed syllable or an unstressable reduced syllable. In the latter case it is unclear what kind of material would be preferred, according to parallelism, between head noun and attribute. What is more, the foot structures of the attributes with initial stress was variable, i.e. some items had initial monosyllabic feet (e.g. *(Haupt)(schu.le)*), some presented with trochaic initials (e.g. *(Ei.sen)(bahn)*).

einer was collapsed with the likewise trochaic prepositional *von der* and juxtaposed to the monosyllabic genitive determiner *der*.

Table 3 shows the percentages of monosyllabic (*der*) vs. disyllabic responses (*einer* or *von der*) broken down by the prosodic status (stressed or unstressed) of the ultima of the head noun and the initial syllable of the attribute noun. Clearly, participants gave more disyllabic responses when the head noun presented with stress on the ultima and when the attribute had initial stress.

Logistic mixed models (Bates et al. 2013) were applied to assess the effects of the prosodic status of the head noun (stressed or unstressed ultima) as well as of the attribute noun (stressed or unstressed initial syllable) on the choice of monosyllabic or disyllabic responses. The intercepts for participants and items were included as random effects. Again, as in the previous study, predictor variables (which were coded as orthogonal sum contrasts) were culled from the model when their inclusion did not improve model fit.

The results of the best fitting logistic mixed model are tabulated in Table 4. Contrary to predictions, including the effect of stress position on the attribute does not improve model fit. However, the model confirms that the prosodic structure of the head noun significantly affects the choice of the construction. With an unstressed ultima on the head noun, the monosyllabic determiner is clearly preferred over the disyllabic genitive or prepositional construction, most likely because the latter would yield a disrhythmic structure with three or four unaccented syllables in a row. The results thus support the hypothesis that participants strive for rhythmic alternation when making syntactic decisions.[11]

There is, however, no indication that participants build prosodically parallel structures. According to prosodic parallelism, as conceived by Wiese and Speyer (2015), participants would have had to prefer a disyllabic trochaic genitive or preposition plus determiner after a trochaic head noun, or, conversely, a monosyllabic genitive after a head noun featuring a stressed ultima. This is clearly not the case.

To conclude, the propensity for rhythmic alternation has a significant impact on construction choice while any effect of prosodic parallelism remains mute.

11 In recent years, a number of online experiments studying eye movements in reading showed comparable results which suggest that the rhythmic/prosodic environment affects syntactic parsing decisions in written sentence comprehension (Breen and Clifton 2013; Kentner 2012; Kentner and Vasishth 2016).

Tab. 4: Coefficients of the best fitting generalised linear model evaluating the choice of the possessive/partitive construction

	Estimate	Std. Error	z-value	p-value
(Intercept)	−0.33166	0.14641	−2.265	0.0235
UltimaHeadNoun	0.09413	0.03711	2.536	0.0112

2.3 *Clash, *Lapse, Parallelism – a corpus study

The third study examines a large-scale corpus (DeReKo, cf. Institut für Deutsche Sprache [IDS]) to directly compare the effects of *Clash, *Lapse, and Parallelism on the morphophonological variation concerning the German adverbs *gern ~ gerne* ('happily'), *lang ~ lange* ('for long'), *selbst ~ selber* ('{my-, your-, her-, him-, our-, them}-{self-, -ves}'), and *meist ~ meistens* ('most of the time'). These adverbs display a (free) alternation concerning the schwa and, consequentially, syllabic structure, i.e. they feature either a monosyllable or a trochee. In the latter case, the final syllable is always a reduced syllable (schwa or [ɐ] in the case of *selber*). Importantly, in contrast to further alternating adverbs, the allomorphs of these adverbs have graphemic cognates both of which are equally acceptable in written Standard German. To the best of my knowledge, there are no more alternating adverbs both variants of which are likewise acceptable in writing.

2.3.1 Method and materials

The frequencies of the four variable adverbs were examined in the context of two forms of the verbs *tun* and *machen* ('to do', 'to make') when these follow the variable adverb. This way, four variable adverbs by two verb forms, i.e., eight quadruplets of prosodically different adverb-verb combinations were scrutinised.

Tab. 5: Bigrams scrutinised in corpus experiment and corresponding factors used for the evaluation of the rhythmic effects

Adverb	Verb	*CLASH	*LAPSE	PARALLEL
gern/selbst/meist/lang	*tun*	×	✓	✓
gerne/selber/meistens/lange	*tun*	✓	✓	×
gern/selbst/meist/lang	*getan*	✓	✓	×
gerne/selber/meistens/lange	*getan*	✓	×	×
gern/selbst/meist/lang	*machen*	×	✓	×
gerne/selber/meistens/lange	*machen*	✓	✓	✓
gern/selbst/meist/lang	*gemacht*	✓	✓	×
gerne/selber/meistens/lange	*gemacht*	✓	×	×

The prosodic profile of each bigram was coded according to the three rhythmic constraints. This was done in a binary fashion, as displayed in Table 5, where the bigrams are represented as either respecting or violating each of the three constraints respectively.

For each of the four combinations of verb form and adverb, the bigram frequencies within the DeReKo corpus, written section (Institut für Deutsche Sprache [IDS]) were determined. Chi-square tests were applied to test the statistical independence of adverb and verb form. These tests use contingency tables like (14) to compare the expected frequencies according to the null hypothesis (which assumes adverb and verb forms to be statistically independent from each other) to the actual, observed frequencies.

(14)

		Prosodic form of verb	
		Monosyllabic	**iambic**
Prosodic form of adverb	monosyllabic	*gern tun*	*gern getan*
	trochaic	*gerne tun*	*gerne getan*

2.3.2 Data analysis and results

For seven of the eight quadruplets of bigrams, the Chi-square tests clearly disprove statistical independence of the prosodic structure of adverb and verb (with p-values < 0.01), supporting the hypothesis that the choice of the prosodic form is conditioned by the prosodic shape of the context. Only in the case of the *meist(ens) machen/gemacht* quadruplet, the test did not yield a significant result. In any case, it has to be determined whether and to what extent each of the three rhythmic constraints under discussion contribute to the prosodic effect. Therefore, for each of the 32 bigrams, the standardised Chi square residuals[12] were calculated as a measure for the degree of deviance from assumed statistical independence of the prosodic form of the adverb and the prosodic form of the verb. Testing the predictions of the three rhythmic constraints against the residuals can inform us about the extent to which each constraint contributes to the frequency distribution of the adverb-verb combinations. In general, a negative residual indicates that a bigram occurs less frequently than the null hypothesis would lead one to expect; conversely, a positive value indicates that the bigram is used more frequently than expected. That is, if the constraints were to affect the prosodic form of the adverb-verb bigrams, structures that violate a given constraint should obtain negative residuals, while bigrams that respect the constraint should engender more positive residuals.

12 Standardised residuals are calculated as (Observed Frequency – Expected Frequency) / sqrt(Expected Frequency)

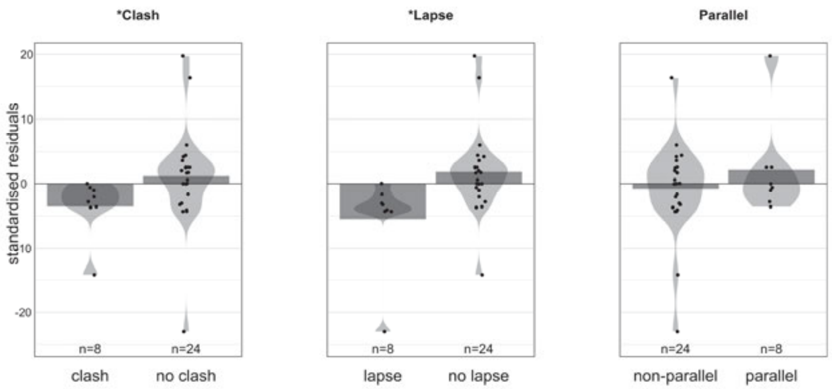


Fig. 1: Standardised Chi square residuals (y-axis) broken down by the two levels of each factor, representing the constraints *CLASH (left panel), *LAPSE (middle panel) and PARALLELISM (right panel), respectively (x-axis). The dots correspond to the individual residuals for each bigram (n=32). The bars represent the mean residual, and the shaded area around the dots illustrates the density of the distribution (the wider the shaded area, the denser the clustering of the residuals in that area).

In order to get a first impression about the contribution of the three constraints, the 32 standardised residuals are plotted for each level of the three predictor variables using the YaRrr package (Phillips 2017) in the statistical computing environment R (R Core Team 2015). As the plot in Figure 1 shows, bigrams that violate a constraint are, on average, less frequent than expected according to the null hypothesis and hence show more negative residuals (cf. left bars of the three panels) while bigrams respecting the constraints are more frequent than expected. The distribution of residuals thus corroborates the hypothesis that adverb-verb bigrams that respect the rhythmic constraints are favored over those bigrams that violate the relevant constraints. However, the differences between the residuals for the bigrams that violate versus bigrams that obey a given constraint are clearly more pronounced in the case of *Clash and *Lapse than in the case of Parallelism. This is especially apparent in the residuals for those bigrams that violate *Clash and *Lapse: Almost all residuals for bigrams that involve a clash or a lapse are negative, while the residuals for the non-parallel bigrams (left bar in the right panel of Fig. 1) are more evenly distributed with the mean residual close to zero. The distribution of residuals corresponding to the bigrams respecting *Clash or *Lapse (right bars in the left and middle panel), while positive on average, spans both the positive and the negative range (most likely due to the fact that bigrams that obey *Clash may violate *Lapse, and vice versa).

Linear models (Bates et al. 2013) were employed to analyse the data. The standardised residuals that were calculated for each of the 32 bigrams (see above) were used as dependent variable. The three constraints (*Clash, *Lapse, Parallelism) served as binary predictor variables, with each bigram violating or respecting the constraints (cf. Table 3); these predictors were coded as orthogonal contrasts. Including the specific adverb as grouping variable (random effect) did not improve model fit. In Table 6, the output of the model including all three predictor variables is tabulated, with *Clash and *Lapse clearly showing significant effects while the effect of Parallelism remains non-significant.

A second, simpler model was fit with Parallelism discarded as predictor (cf. Table 7). Applying the anova function to compare the simpler model with the full model suggests that discarding Parallelism does not deteriorate model fit ($Df = 1$, $p = 0.32$).

To summarize, the negative Chi square residuals for bigrams involving a stress clash (e.g. *gern machen*) or a stress lapse (e.g. *gerne getan*) reflect the avoidance of these rhythmically sub-optimal structures when compared to bigrams that obey the respective constraints. No such pattern of avoidance could be observed for bigrams that violate the Parallelism constraint (i.e. non-parallel bigrams like *gerne tun* or *gern machen*). This corpus study thus corroborates the hypothesis that the inclusion or omission of the optional schwa-syllable on the adverb is conditioned by the stress status of the initial syllable of the verb. The overall prosodic shape of the verb, however, i.e. whether it is monosyllabic, trochaic, or iambic, does not appear to affect the inclusion/omission of the schwa syllable on the adverb beyond the effects of *Clash and *Lapse.

Tab. 6: Model including all three main effects

	Estimate	Std. Error	t-value	p-value
(Intercept)	–0.02698	1.02653	–0.026	0.979212
*Clash	2.13	0.64902	3.282	0.00277
*Lapse	2.29243	0.64902	3.532	0.00145
Parallelism	0.65506	0.64902	1.009	0.32148

Tab. 7: Model with main effect of PARALLELISM culled

	Estimate	Std. Error	t-value	p-value
(Intercept)	−0.02698	1.02653	−0.026	0.979212
*Clash	1.96624	0.62862	3.128	0.003987
*Lapse	2.45619	0.62862	3.907	0.000514

2.3.3 Discussion

This corpus study yields important insights regarding the morphophonological variation on the adverbs under study. First of all, provided that the written corpus does in fact reflect prosodic preferences, it is clear from the results that supralexical prosodic structure co-determines the presence or absence of a reduced syllable on the variable adverbs. This is in accordance with similar findings by Ingason (2015), Kaufmann (2014), Schlüter (2005), and Vogel et al. (2015) who report rhythmic influences on morphological or morphosyntactic variation. Secondly, this study fails to replicate the findings by Wiese and Speyer (2015) who hold prosodic parallelism accountable for the presence or absence of a reduced syllable. In this study, PARALLELISM does not appear to contribute to the morphophonological variation of the adverbs. The model comparison suggests that the rhythmic influences are reducible to *CLASH and *LAPSE alone. One conceivable reason for the discrepancy between the present results and the findings by Wiese and Speyer (2015) lies in the difference between the structures scrutinised: while this study looked at prosodically variable adverb-verb sequences (e.g. *gern(e) tun*), Wiese and Speyer (2015) studied the variable adverb in other contexts (e.g. the verb-adverb sequence *wär(e) gern(e)*). It remains to be seen why prosodic parallelism explains the variation in one case but not in the other. In this context, it would also be interesting to check to what extend the rhythmic constraints *CLASH and *LAPSE contribute to the variance in Wiese and Speyer's dataset.

Furthermore, this study reveals an interesting finding regarding the relative contributions of *CLASH and *LAPSE, with the latter apparently having a similar, if not stronger, impact on morphophonological choice when compared to *CLASH. Given the greater attention to stress clash and its avoidance in the literature and the comparatively limited consideration of the *LAPSE constraint, this may seem astonishing (entering the terms 'clash' and 'lapse' in the context of the phrase 'linguistic rhythm' produces 493 hits for 'clash' but only 271 for 'lapse' on Google Scholar). What is more, as noted by Julia Schlüter,

> [...] many authors [...] concur in the view that stress clashes are perceived as far more objectionable than stress lapses; while the latter are tolerated to a certain extent, the former almost categorically necessitate compensatory measures.
>
> (Schlüter 2005: 20)

Possibly, the somewhat weaker effect of *CLASH on presence or absence of schwa is due to the fact that a stress clash may be alleviated in other ways, e.g. by stress retraction or stress promotion, processes that the writer may subconsciously execute (remember that we are dealing with data from a written corpus). Conversely, it is hardly possible to change a structure violating *LAPSE by altering the assignment of prominences to syllables because the unstressable reduced syllables simply cannot become stressed. A writer abiding by the principle of rhythmic alternation is thus more likely to put morphophonological variation to its rhythmic use in the event of a potential lapse than in the event of a potential clash (see Shih et al. 2015, for a similar point).

In the following, I note several limitations of this study. For one thing, since I examined the variable structures within a written corpus only, it remains unclear whether the results are generalisable to the oral modality. Even more importantly, since only bigrams were studied, with the wider (prosodic) context disregarded, the validity of the results is open to suspicion. It is quite possible that an analysis that considers the phrasal context would lead to different results. However, the approach taken here is in keeping with Wiese and Speyer (2015) who also only considered bigrams, rendering the studies at least methodologically comparable. Finally, the scope of this study is very narrow, narrower by far compared to Wiese and Speyer (2015) who consider schwa-zero variation not only on adverbs but in many more contexts. The results therefore have to be taken with some caution.

3 General discussion and conclusion

Overall, the three studies presented here clearly support the claim that the rhythmic-prosodic context affects morphophonological variation. The first study revealed an effect of the rhythmic pattern (due to the distribution of lexical stresses) on the realisation of the variable adverb *gern(e)* in oral reading. The second study, a forced choice experiment, showed that the variable morphosyntax of the possessive or partitive relation is susceptible to rhythmic structure.

Finally, a corpus study demonstrates the non-independence of the prosodic shapes of variable adverbs in adverb-verb sequences.

As to the relative contribution of the three constraints under discussion (*Clash, *Lapse, Parallelism) for explaining the variance observed in the three experiments, the findings paint a somewhat mixed picture. The final corpus study quite clearly dismisses the importance of Parallelism, while showing that *Lapse and *Clash, have a clear impact on the choice of monosyllabic vs. trochaic adverb. Similarly, the experiment on the choice between morphological genitive and prepositional phrase reveals a weak effect of rhythmic alternation but fails to reveal an effect of prosodic parallelism.

The first experiment, however, suggests that prosodic parallelism has a role to play in the realisation of the variable adverb in oral reading. It shows that an iterating rhythm is effectively priming the morphophonological form of the variable adverb that continues the preceding (trochaic) rhythm. However, as highlighted in the discussion of that experiment, the iterating rhythm is only observable through the lens of certain assumptions regarding the foot structure involved, i.e. it is only valid when foot boundaries are allowed to straddle word boundaries (contra the Strict Layer Hypothesis) and when adverbs may be demoted to a prosodically weak position. That is, while there is clear evidence for the joint effects of *Clash and *Lapse conditioning the morphophonological structure of words and phrases, effects of prosodic parallelism are relatively minute. This is not to contest the relevance for prosodic parallelism in other contexts. As discussed in the introduction, prosodic parallelism is likely to be a constraining factor in word formation (e.g. reduplication) and it is clearly involved in poetic language. Quite possibly, the role of Parallelism is more pronounced in more artistic language use or, more generally in circumstances that are not as strictly constrained by time. Note that for Parallelism to become apparent, the linguistic processor needs to consider more material (at least two adjacent feet) than when evaluating local rhythmic well-formedness on a syllable-to-syllable basis.

All in all, the results of the studies presented suggest that phrases and sentences are not built by merely concatenating morphs according to a pre-specified syntactic structure. In addition, word forms may be altered in various ways to suit the supra-lexical rhythmic structure, and the rhythmic structure may reciprocally co-determine morphosyntactic choice.

Acknowledgements: I thank the organisers and the audience of the workshop on empirical approaches to the phonological structure of words (Marburg, October 2014) for the opportunity to present and discuss this work. Caroline Féry, Isabelle Franz, Ingo Plag, Renate Raffelsiefen and two anonymous reviewers provided helpful comments and constructive feedback on the manuscript.

References

Anttila, Arto, Matthew Adams & Michael Speriosu. 2010. The role of prosody in the English dative alternation. Language and Cognitive Processes 25. 946–981.

Bates, Douglas, Martin Maechler, Ben Bolker & Steven Walker. 2013. lme4: Linear mixed-effects models using eigen and s4. R package version 1. https://CRAN.R-project.org/package=lme4 (accessed 12 January 2015).

Behaghel, Otto. 1930. Zur Wortstellung des Deutschen. Language 6(4). 29–33.

Breen, Mara. 2014. Empirical investigations of the role of implicit prosody in sentence processing. Language and Linguistics Compass 8. 37–50.

Breen, Mara & Charles Clifton Jr. 2013. Stress matters revisited: A boundary change experiment. The Quarterly Journal of Experimental Psychology 66. 1896–1909.

Casteren, Maarten van & Matt Davis. 2006. Mix, a program for pseudorandomization. Behavior Research Methods 38. 584–589.

Chafe, Wallace. 1988. Punctuation and the prosody of written language. Written Communication 5. 395–426.

Domahs, Ulrike, Frank Domahs & Christina Kauschke. 2018. The morphology-prosody interface in typically developing and language-impaired populations: An overview and insights from experimental approaches. In Christiane Ulbrich, Alexander Werth & Richard Wiese (eds.), Empirical approaches to the phonological structure of words (Linguistische Arbeiten 567), 95–120. Berlin & Boston: De Gruyter.

Duanmu, San. 2012. Word-length preferences in Chinese: a corpus study. Journal of East Asian Linguistics 21. 89–114.

Eisenberg, Peter. 1991. Syllabische Struktur und Wortakzent: Prinzipien der Prosodik deutscher Wörter. Zeitschrift für Sprachwissenschaft 10. 37–64.

Evertz, Martin & Beatrice Primus. 2013. The graphematic foot in English and German. Writing Systems Research 5. 1–23.

Fanselow, Gisbert & Caroline Féry. 2002. Ineffability in grammar. In Gisbert Fanselow & Caroline Féry (eds.), Resolving conflicts in grammars: Optimality theory in syntax, morphology, and phonology, 265–307. Hamburg: Helmut Buske Verlag.

Fehringer, Carol. 2011. Allomorphy in the German genitive: A paradigmatic account. Zeitschrift für Germanistische Linguistik 39. 90–112.

Féry, Caroline. 1994. Umlaut and Inflection in German. ROA – Rutgers Optimality Archive 34. 1–44.

Féry, Caroline. 1997. Uni und Studis: die besten Wörter des Deutschen. Linguistische Berichte 172. 461–489.

Féry, Caroline & Gerrit Kentner. 2010. The prosody of embedded coordinations in German and Hindi. In Proceedings of Speech Prosody, vol. 5. Chicago.

Ghini, Mirco. 1993. ϕ-formation in Italian: A new proposal. Toronto Working Papers in Linguistics, 41–78.
Gutzmann, Daniel & Katharina Turgay. 2015. Expressive intensifiers and external degree modification. The Journal of Comparative Germanic Linguistics 17. 185–228.
Ingason, Anton Karl. 2015. Rhythmic preferences in morphosyntactic variation and the theory of loser candidates. In Ralf Vogel & Ruben van de Vijver (eds.), Rhythm in Cognition and Grammar: A Germanic Perspective, 235–253. Berlin: De Gruyter.
Institut für Deutsche Sprache (IDS). 2014. Deutsches Referenzkorpus. Archiv der Deutschen Gegenwartssprache [Computer Software]. https://cosmas2.ids-mannheim.de/cosmas2-web/ (accessed 12 January 2015).
Itô, Junko & Armin Mester. 1997. Sympathy theory and German truncations. University of Maryland Working Papers in Linguistics 5. 117–139.
Kager, René 1989. A metrical theory of stress and destressing in English and Dutch. Dordrecht: Foris.
Kallulli, Dalina & Antonia Rothmayr. 2008. The syntax and semantics of indefinite determiner doubling constructions in varieties of German. The Journal of Comparative Germanic Linguistics 11. 95–136.
Kaufmann, Göz. 2014. Die Tür(e) in der deutschen Literatur – Tor zum Verständnis sprachlicher Variation? In Pia Bergmann, Karin Birkner, Peter Gilles, Helmut Spiekermann & Tobias Streck (eds.), Sprache im Gebrauch: räumlich, zeitlich, interaktional. Festschrift für Peter Auer, 123–142. Heidelberg: Winter.
Kentner, Gerrit. 2012. Linguistic rhythm guides parsing decisions in written sentence comprehension. Cognition 123. 1–20.
Kentner, Gerrit. 2015. Problems of prosodic parallelism: A reply to Wiese & Speyer (2015). Linguistics 53. 1233–1241.
Kentner, Gerrit. 2017. On the emergence of reduplication in German morphophonology. Zeitschrift für Sprachwissenschaft 36. 233–277.
Kentner, Gerrit & Caroline Féry. 2013. A new approach to prosodic grouping. The Linguistic Review 30. 277–311.
Kentner, Gerrit & Karsten Steinhauer. 2017. Prosody in written language – acquisition, processing, emergence of norms. Lausanne: Frontiers Media SA.
Kentner, Gerrit & Shravan Vasishth. 2016. Prosodic focus marking in silent reading: effects of discourse context and rhythm. Frontiers in Psychology 7. 319.
Köpcke, Klaus-Michael. 2002. Die sogenannte i-Derivation in der deutschen Gegenwartssprache: Ein Fall für outputorientierte Wortbildung. Zeitschrift für germanistische Linguistik 30. 293–309.
Kuijpers, Cecile & Wilma van Donselaar. 1998. The influence of rhythmic context on schwa epenthesis and schwa deletion in Dutch. Language and Speech 41. 87–108.
Kutscher, Anna. 2016. Optimale Adverbpositionen an der Schnittstelle zwischen Semantik, Phonologie und Syntax. Bielefeld: Universität Bielefeld.
Leiner, Dominik J. 2014. Sosci Survey [Computer Software]. https://soscisurvey.de/ (accessed 12 January 2015).
Luo, Yingyi, Yunyan Duan & Xiaolin Zhou. 2015. Processing rhythmic pattern during Chinese sentence reading: An eye movement study. Frontiers in Psychology: Language Science 6. 1881.
Luo, Yingyi & Xiaolin Zhou. 2010. ERP evidence for the online processing of rhythmic pattern during Chinese sentence reading. NeuroImage 49. 2836–2849.

Menninghaus, Winfried, Valentin Wagner, Eugen Wassiliwizky, Thomas Jacobsen & Christine A. Knoop. 2017. The emotional and aesthetic powers of parallelistic diction. Poetics 63. 47–59.
Myrberg, Sara. 2013. Sisterhood in prosodic branching. Phonology 30. 73–124.
Phillips, Nathaniel. 2017. YaRrr!: The Pirate's Guide to R. http://CRAN.R-project.org/package=yarrr (accessed 07 June 2017).
R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://R-project.org (accessed 12 January 2016).
Rohdenburg, Günter. 2014. Grammatical variation and the avoidance of stress clashes in Northern Low German. In Jenny Cheshire & Dieter Stein (eds.), Taming the vernacular: From dialect to written standard language, 93–109. New York: Routledge.
Schlüter, Julia. 2005. Rhythmic grammar: the influence of rhythm on grammatical variation and change in English. Berlin: Mouton de Gruyter.
Selkirk, Elizabeth. 1984. Phonology and syntax. Cambridge, Mass.: MIT Press.
Selkirk, Elizabeth. 2000. The interactions of constraints on prosodic phrasing. In Merle Horne (ed.), Prosody: Theory and experiment, 231–261. Dordrecht: Kluwer Academic.
Shih, Stephanie, Jason Grafmiller, Richard Futrell & Joan Bresnan. 2015. Rhythm's role in genitive construction choice in spoken English. In Ralf Vogel & Ruben van de Vijver (eds.), Rhythm in Cognition and Grammar: A Germanic Perspective, 207–233. Berlin: De Gruyter.
Speyer, Augustin. 2010. Topicalization and stress clash avoidance in the history of English. Berlin: Mouton de Gruyter.
Vogel, Ralf, Ruben van de Vijver, Sonja Kotz, Anna Kutscher & Petra Wagner. 2015. Function words in rhythmic optimisation. In Ralf Vogel & Ruben van de Vijver (eds.), Rhythm in Cognition and Grammar: A Germanic Perspective, 255–276. Berlin: De Gruyter.
Wegener, Heide. 2004. Pizzas und Pizzen – die Pluralformen (un)assimilierter Fremdwörter im Deutschen. Zeitschrift für Sprachwissenschaft 23. 47–112.
Wiese, Richard. 2009. The grammar and typology of plural noun inflection in varieties of German. The Journal of Comparative Germanic Linguistics 12. 137–173.
Wiese, Richard. 2016. Prosodic parallelism – comparing spoken and written language. Frontiers in Psychology 7. 1598.
Wiese, Richard & Augustin Speyer. 2015. Prosodic parallelism explaining morphophonological variation in German. Linguistics 53. 525–559.

Christiane Ulbrich and Richard Wiese

Phonotactic principles and exposure in second language processing

Abstract: The processing of phonotactic patterns is crucial in any language for the recognition of word boundaries in running speech. The present paper reports the results of a reaction time experiment with native Russian and Chinese L2 learners of German. Chinese and Russian differ from German in the complexity of possible sound combinations in the syllable coda, in that Russian allows more consonant sequences and more frequent violation of sonority requirements compared to German, whereas Chinese does not allow complex consonant clusters in final syllable position. Applying an artificial language paradigm, we investigate the role of language-specific requirements of sound sequences in the syllable coda for the processing of L2 phonotactics and the influence of exposure to more or less restrictions on the sonority requirements. In a word-picture matching task, participants were exposed to nonce words with existent and non-existent German final consonant clusters violating sonority requirements or not. Reaction time data were obtained in the recollection of word-picture pairs. The results show that recollection and processing of final consonant clusters is facilitated by both adherence to sonority requirements and the exposure to existing consonant clusters. However, the influence of implicit knowledge of universal phonotactic principles and frequency-based factors varies depending on phonotactic properties of the native language.

Keywords: phonotactics, second language, sonority

1 Introduction

One of the central questions in research on second language acquisition (SLA) is the nature of interference between the first (L1) and the second language (L2). For several decades, a general consensus exists that language-specific characteristics of an L1 transfer into the production and perception of an L2 (e.g. Odlin 1989; Ellis

Christiane Ulbrich: Universität Konstanz, Universitätsstraße 10, D-78464 Konstanz, christiane.ulbrich@uni-konstanz.de
Richard Wiese: Philipps-Universität Marburg, Pilgrimstein 16, D-35032 Marburg, wiese@uni-marburg.de

DOI 10.1515/9783110542899-007

1994; De Bot 1986; Grosjean 1998; Major 1994; Benson 2002; Ulbrich 2013). In addition, more recently, extensive work in the field has produced ample evidence for an impact of L2 on L1 (e.g. Pavlenko and Jarvis 2002; Pavlenko 2000; Ulbrich and Ordin 2014; Mennen 2004; Flege 2002; Cook 2003; Chang 2010). However, it remains an unresolved issue which aspects of the phonological knowledge interact in the process of language learning. To complicate the matter, input factors such as experience, frequency and exposure have frequently been shown to facilitate or hinder language mastery. Hence both linguistic and non-linguistic factors influence the process of SLA, but it is unclear what their relative impact is in phonological processing.

In order to explain and model SLA, two opposing theoretical approaches currently compete. On the one hand, theories of universal grammar (UG) attempt to explain acquisition of a second (or any additional) language on the basis of underlying universal principles. The phonological system of a language user is based on abstract rules, constraints, or principles which are categorical and generalized across languages. On the other hand, usage-based approaches understand language and its regulations as a dynamic adaptive system (Ellis and Larsen-Freeman 2006). Only general cognitive functions and language use are the prerequisites for emergent categorisation and generalisations (Bybee and Hopper 2001). Language use is determined by exposure and frequency in the input. Thus, a comprehensive theory of SLA must supply a theoretical base allowing for integration of neuro-cognitive and environmental factors which engender linguistic behavior. However, it does not seem inconceivable that both approaches apply in the process of language learning, i.e., that language use is grounded in both abstract principles and input patterns (Moisik 2009; Ellis 2005).

2 Phonotactics in L2 phonology

During the first year of monolingual first language acquisition, a children's ability to process and produce all possible speech sounds of any world's language weakens because the L1 system develops according to the input of the ambient language. During acquisition and language mastery children become more and more constrained by the systematic organization of the L1, referred to as a phonological filter of the L1. Thus, perception is shaped by the L1 perceptual system (Strange 1995: 22 and 39). Furthermore, a number of studies support the application of a phonological filter in language processing. For instance, studies by

Domahs et al. 2009; Berent and Lennertz 2010; Berent et al. 2014 show that monolingual individuals perceive and interpret words and nonce words on the basis of their native language.

The implication for subsequent L2 acquisition is that the phonological rules of the L1 – including language-specific phonotactic regularities – influence processing and production of L2 phonology (see Boll-Avetisyan 2018, in this volume). The influence of the L1 is evidenced by a number of perception studies. Halle et al. (1998), for example, report perceptual assimilation in numerous tasks, in which speakers tend to mispronounce, substitute and adapt L2 consonant clusters illegal in their L1 to clusters which are legal in their L1.

The assumption of an L1 filter is not new. Already Trubetzkoy (1939/1967) assumed a perceptual base for erroneous or deviant L2 sound production and suggested the L1 system to act as a 'phonological filter' through which L2 sounds are perceived and ultimately classified. However, in the current debate there is disagreement as to whether this L1 impact holds for all stages of L2 development. Some believe this to be true (e.g. Bley-Vroman 1989; Brown 1998), while others convincingly argue for the sensitivity of L2 learners to phonotactic regularities of the target language (Cook 1991; Schwartz and Sprouse 1996; Halle et al. 1998; van Heuven, Dijkstra, and Grainger 1998; Jared and Kroll 2001). In a lexical decision task with English native speakers and sequential Russian-English bilinguals, Mikhaylova (2009) reports differences between the two subject groups, in that bilinguals were overall slower and less accurate in their performance of the task. However, the overall pattern of the results was the same: legal nonce words are processed faster and recognised as possible words more accurately than illegal nonce words. Phonotactically legal real words were identified as legal nonce words which Mikhaylova (2009) explains by the fact that participants relied both on their lexical knowledge and the high frequency and neighbourhood density of the real items. The similarity between the native and the non-native subject group in response to English phonotactic constraints indicates that the L2 speakers access phonotactic constraints that cannot be transferred from L1 (see also van Heuven, Dijkstra, and Grainger 1998; Jared and Kroll 2001). This in turn questions the immutability and the strength of an L1 phonological filter and suggests that with increasing language mastery the L1 filter can be overcome, and the developing phonological system of the interim language can include L2 constraints. For the study to be presented below it is relevant that this sensitivity increases with language proficiency.

In another study by Halicki (2010), native-like intuitions of word transformations in L2 are tested in English adult L2 learners of French. The study investigates the knowledge of L2 phonotactics on the basis of consonant clusters (CC and CCC). The results show that explicit teaching and L1 transfer facilitate well-formedness judgements of French words performed by intermediate L2 learners of French. It remains unclear what the source of the learner knowledge may be. Due to the relative infrequent occurrence of the tested items in French (CCC and r+k clusters), frequency-based learning only cannot explain the results, and the author concludes in favour of an L2 phonological grammar that is also constrained by UG principles. This conclusion is in line with other proposals of an L2 phonology in which UG and input factors interact in the development (e.g., Archibald 2004; Dresher 1999).

The present paper contributes to the debate on the influence of underlying universal constraints and exposure on the acquisition of a second language. Some phonotactic constraints are assumed to be part of universal linguistic, cognitive, articulatory, or auditory systems because of their cross-linguistic similarity (e.g., Moreton 2002). However, languages differ considerably in their representation of phonotactic constraints, so that it seems equally necessary to assume that at least parts of this phonological knowledge have to be learned. We are interested in the question whether or not the language-specific setting of a phonotactic constraint in the L1 changes as a result of advancing L2 mastery. In other words, the study contributes to the understanding of the nature and strength of a phonological filter. The paper discusses these issues by investigating processing speed and accuracy in the recollection of nonce words in a word-picture matching task performed by Russian and Chinese L2 speakers of German. On the basis of the analysis of reaction times and accuracy in a recollection task of word-picture pairs, the aim of the study was to test the interplay between phonotactic regularities and input factors in second language learning.

3 Phonotactics and sonority

Syllables consist of a sequence of segments or other units. However, phonotactic rules and constraints that govern the linear distribution of sounds differ across languages. One view holds that the sonority hierarchy universally determines the structure of syllables, where sonority is the assumed property of "inherent loudness", which is lowest for stops and highest for vowels. The Sonority Sequencing Principle (SSP) entails that elements that are more sonorous are closer to the syllable nucleus, while less sonorous elements are closer to the syllable edges. For example, an English or German syllable spelt *blank* has two minimally sonorous sounds at the left and right edge (/b, k/), outside of two sounds with intermediate sonority values (/l, ŋ/), and the most sonorous vowel /æ/ or /a/, respectively, in the nucleus or peak of the syllable.

The discussion of phonotactic principles has a long tradition in linguistic inquiries (e.g. Whitney 1865; Sievers 1901; for a recent survey, see Parker 2012), and it has been shown that the orderly sequencing of phonemes facilitates the identification, segmentation and processing of syllables and words. However, the sonority model of the syllable has always been under debate (see summaries in Parker 2012, and Wiese 2016); various definitions of the sonority scale (using articulatory, acoustic, or perceptual criteria) have been proposed, and numerous exceptions to the sonority generalization exist. The present paper attempts, inter alia, to study the role of sonority-related patterns in the processing of language, with the most commonly used definition of sonority based on articulatory features.

For this purpose, the sonority scale given in (1) is assumed. It ranks classes of segments according to the degree of opening in the vocal tract required for the articulation of sounds of the respective class. ‘<’ is an abbreviation for ‘less sonorous than’. We also give an example from each class.

(1) Sonority hierarchy

plosive < affricate < fricative < nasal < liquid < glide < vowel

/b/ /ts/ /f/ /n/ /l/ /j/ /i/

This scale distinguishes five degrees of sonority for consonants and thus allows for a fine-grained distinction within consonant clusters. For example, a final cluster /fts/ with a fricative segment followed by the affricate /ts/ would be one adhering to the sonority hierarchy specified in (1), while a cluster /tsf/ would exemplify a violating cluster. While sonority as defined here is only one of the principles discussed for syllable phonology and is complemented by others such as identity avoidance (Obligatory Contour Principle, OCP), preference for CV syllables over all other types such as VC or CCV, it has certainly been considered both as a central principle as well as a debated one.

German is a language allowing reasonably complex consonantal clusters word-initially and word-finally. In word-final position, up to four consonants can be found, as in *Herbst* ‘autumn’. Constraints for these combinations have been discussed in terms of cluster size (clusters of length 2 are more common than longer ones), place features (tongue-tip sounds are more common), and sonority; see Vennemann (1972), Wiese (1988) or Hall (1992) for a more detailed discussion.

Explanations for the fact that some combinations exist, while many others do not, have been sought in terms of phonological markedness and/or perceptual preferences.

In any case, it is obvious that 20 to 25 consonantal phonemes in German (depending on the analysis of complex and/or marginal phonemes) yield only about 50 to 54 bi-segmental word-final combinations (Wiese 2000), with the exact number depending on the treatment of marginal clusters found in loan words. That is, complex clusters can mark the right word-edge, but there are many gaps, some of which do not appear to be accidental.

This scenario makes it possible to study the processing of various types of consonant clusters. In the following, the focus is, first, on the distinction between existent (EX) and non-existent (NX) clusters, and, second, on the distinction between well-formed (WF) and ill-formed (IF) clusters.

Well-formedness is defined for the present study by the notion of sonority. Crucially, both types of clusters exist in languages such as German, and the existence of clusters violating the sonority principle may be seen either as evidence against this principle or as counterexamples to an otherwise valid principle (Parker 2012). In any case, the fact that "illegal" clusters exist will be used in the present study.

As noted, languages may vary in the restrictions they place on the clustering of consonants. Standard Chinese (Duanmu 2002, 2014) is a language with a maximal syllable template of CCVC (Wiese 1997). The second C can only be filled with a glide, and the syllable template allows only one syllable-final nasal consonant, as in *tan* or *ping* (/tan/, /pɪŋ/). Even a post-vocalic glide as in *lai* (/laɪ̯/) prevents the occurrence of any additional consonant. Thus, Chinese exemplifies a language in which speakers have only very limited exposure to complex phonotactic patterns. A hierarchy referring to consonants, glides and full vowels is sufficient to describe syllabic patterning in this language.

On the other hand, Russian shows a more extended patterning of consonant-clusters than German (Timberlake 2004; Davidson and Roon 2008). For example, while German allows a limited set of triconsonantal sequences word-initially, Russian displays a wide range of sequences with up to four consonants. Word-finally, there are again few restrictions, so that even many obstruent-obstruent clusters are admitted. While most studies of Russian phonology avoid the topic of consonant clusters altogether, the comprehensive Grammar of Russian by Švedova (1980) contains long lists of bisegmental, trisegmental and quadrisegmental final clusters. In summary, Chinese and Russian contrast with German in terms of phonotactics: While Chinese speakers have been exposed to a very limited amount of phonotactic patterns with no consonantal clusters, Russian speakers have experienced a wider range of phonotactic patterning than have German speakers. The clusters in Russian include many clusters which are ill-formed with respect to the SSP introduced above.

From the above, several predictions can be derived for the two factors of sonority and exposure:

- Word-picture pairs with illegal clusters, i.e. those violating the SSP, should be recognized both by Chinese and Russian learners of German less accurately and slower than legal nonce words (i.e., those without phonotactic violations).
- Word-picture pairs with clusters that exist in German should be recognized more accurately and faster than those with non-existing cluster. However, it is also possible that adherence to sonority and existence both lead to the initiation of lexical access, so that sublexical pre-processing and lexical access take place simultaneously and therefore decelerate and impede processing.
- Russian learners of German will be less sensitive to the SSP due to the frequent occurrence of its violation in their L1. Their performance at the word-picture-matching task will therefore be faster and more accurate than the performance of Chinese learners of German.
- Since we assume the phonological filter of the respective L1 to lose strength during progressing language mastery, we hypothesize the group of Chinese beginners to be less accurate and slower compared to the advanced learner group.

After the conditions for the stimulus creation have been introduced in the sections to follow, Tables 3 and 4 in Section 5 will specify how these hypotheses will be tested.

4 Experiment

In the study of SLA it is important to control for numerous possible confounding influences and interactions of the phonological system with other linguistic and extralinguistic factors, such as grapheme-phoneme relationship, lexical knowledge, word frequency, etc. In order to isolate the influence of the sonority principle in final consonant clusters, we use nonce words rather than real words in our experiment as detailed below (for other applications of artificial language in the study of phonology see e.g. Wilson 2003; Myers and Padgett 2014; for a review see Moreton and Pater 2012; Mikhaylova 2009).

4.1 Participants

66 L2 learners of German were recruited, 44 Chinese learners and 22 Russian learners. The Russian learners (16 females, 6 males), aged between 20 and 27 (average age 24) had lived a minimum of 2 years in Germany. The 44 Chinese learners were allocated into two groups of advanced learners and beginners. 22 advanced learners (14 females, 8 males) were aged between 21 and 24 (average age 21). They too had lived a minimum of 2 years in Germany. The beginners (16 females, 6 males), aged between 19 and 21 (average age 20) had just moved to Germany and had received no more than three month training in German prior to testing. Participants' language proficiency level was tested according to the Language Placement Test following the standard of the Common European Framework (CEFR). Russian and Chinese participants allocated to the advanced groups of participants were tested C2 and Chinese speakers allocated to the beginners' group were tested A1. All participants were right-handed, with normal-to-corrected sight and hearing and were paid a small fee for their participation. The language used for the instructions and questionnaires was German in order to ensure activation of German phonotactic knowledge.

4.2 Stimuli

The dataset was initially designed to allow for a cross-linguistic comparison between Polish and German (see Ulbrich et al. 2016, for German; and Wiese et al. 2017, for Polish) in that segments and phonotactic principles were tested to hold for both languages, explaining the decision for the chosen over other possible clusters. The stimuli used in the previous and the current experiment were monosyllabic nonce words containing the final CC clusters presented in Table 1.

The selection of final consonant clusters was guided by three considerations:

(1) the limited number of existent clusters in German that do not comply with the sonority requirements. We limited the number of clusters to be used to 21 per group so that we only had to duplicate 10 of the 11 non-existent, ill-formed clusters. This ensured that we could avoid too many repetitions of the same cluster in the creation of stimuli.

(2) the fact that in the actual production several consonants become syllabic in final CC-cluster position in German (e.g., the nasal in /sn/ as well as the liquid in /tr/). Such clusters were not included in the experimental stimuli.

(3) the phonetic similarity between non-existent and existent clusters. We matched non-existent and existent clusters in terms of voicing, place and manner of articulation as closely as possible.

We then created artificial word-stimuli with final consonant clusters adhering to the sonority hierarchy or not. Both types of clusters exist in German. There are also missing clusters so that we could group them into four conditions as presented in Table 1. The four groups derive from the crossing of two factors, namely *formedness* (well-formed vs. ill-formed) and *existence* (existent vs. non-existent).

Tab. 1: Groups of clusters used

	existent clusters (EX)	non-existent clusters (NX)
well-formed clusters (WF)	ft, lç, lf, lk, lm, lʃ, mʃ, ms, nç, nt, ʃt, rf, rk, rm, rʃ, rʃ, sk, sp, tst, xt, ŋk	fp, ftp, lŋ, mk, nk, np, nx, ʃk, ʃts, rŋ, stʃ, sts, tʃk, xk[1], xp, , ŋf, ŋp, ŋtʃ, ŋts, ŋx
ill-formed clusters (IF)	çs, fs, ks, kt, kts, pfs, pʃ, ps, pt, ʃs, xs[2]	fʃ, fx, kp, kʃ, ktʃ, pk, pts, px, ʃf, sʃ, tk, tp, tʃf, tʃs, tʃts, tʃx, tsf, tsʃ, tsx, tx, xf

All clusters were used three times to create monosyllabic nonce words with a CVCC structure. The final consonant cluster was preceded by a CV string consisting of only three different consonants and three different vowels, chosen as unmarked and sufficiently distinct from each other in German. The consonants chosen, i.e. /g/, /f/ and /n/, differ in places (velar, alveolar, labial-dental) and manners (plosive, fricative and nasal) of articulation. Selected vowels /a/, /e/ and /o/ are distinct in height, frontness and roundness. The combination of the consonants and vowels lead to nine different CV sequences that potentially could precede the final consonant clusters: /ge/, /ga/, /go/, /fe/, /fa/, /fo/, /ne/, /na/, /no/. In order to create a sufficient number of stimuli per condition, each final consonant cluster was used in three different contexts, as for example in /gektʃ/, /faktʃ/ and /noktʃ/ or /gasp/, /fosp/ and /nesp/. Thus, a total of 252 nonce words were created and used in the current experiment. The experimental conditions were tested in a 2 x 2 factorial design. The two factors had two levels each; formedness (well-formed (WF) and ill-formed (IF)) and existence (existent (EX) and non-existent (NX)). The factors were crossed so that our participants were

1 The affricates /pf/, /ts/ and /tʃ/ were treated as complex, but monophonemic, segments. Instead of /noxk/, the stimulus /noxt/ was erroneously used in the experiment. The respective results were excluded in the analysis, since /xt/ is an existent cluster.

2 In German, only a very limited number of IF clusters exist so that 10 of the 11 identified clusters had to be used twice. The cluster /kt/ is the one used only once.

presented with an equal number of items in four conditions. The number of critical items used was 21 (types) x 2 (EX vs. NX) x 2 (WF vs. IF) x 3 (three CV contexts).

A phonetically trained female native speaker of German from the Berlin area produced each stimulus word at a normal speech rate. Recordings took place in a sound-attenuated cabin at the linguistic laboratory of the University of Marburg. The words were produced in isolation and recorded directly onto a Mac computer using Amadeus Pro, version 2.1. Speech data were digitized at 44.1 kHz with a 16-bit sampling rate (mono format). Some of the clusters were articulatorily rather demanding, so that recordings took place under the supervision of a phonetician who controlled for the clusters' authentic but clear pronunciation and to avoid unnaturally careful pronunciation. A trained phonetician evaluated the individual stimuli with respect to naturalness.

4.3 Phonetic analysis of stimuli

A phonetic analysis of mean fundamental frequency, mean intensity and mean duration of the stimuli produced was carried out in order to exclude potential artefacts that may result from differences in the phonetic realisation of artificial words; see Table 2.[3]

Tab. 2: Phonetic parameters (means, standard deviations) for items in four conditions

nonce word type	F0 (Hz)	Intensity (dB)	Duration (sec)
WF-EX (n=63)	201.81 (±12.8)	51.04 (±5.1)	0.82 (±0.1)
WF-NX (n=63)	200.36 (±15.5)	50.60 (±6.1)	0.94 (±0.12)
IF-EX (n=63)	212.82 (±14.6)	50.48 (±3.7)	0.98 (±0.1)
IF-NX (n=63)	223.14 (±17.1)	50.59 (±4.4)	0.92 (±0.1)

Intensity does not differ significantly between any of the four conditions used in the stimulus material. Fundamental frequency differs significantly between words with WF and IF clusters. Words with IF-EX clusters also differed from those with IF-NX clusters. However, the difference is smaller than 1.5 semitones and

3 This section repeats the analysis presented in Ulbrich et al. (2016), since the same set of stimuli was used.

therefore, according to psycholinguistic studies not reliably discriminable by humans (e.g., Nooteboom 1997; ‘t Hart 1981; ‘t Hart, Collier, and Cohen 1990: 29; Rietveld and Gussenhoven 1985: 304). Significant differences in word length were found between all conditions with the exception of WF-NX and both IF-EX and IF-NX. The differences in duration are 58 ms between IF-EX and IF-NX, 166 ms between WF-EX and IF-EX, 128 ms between WF-EX and WF-NX and 110 ms between WF-EX and IF-NX. However, word length can only be fully evaluated at the end of a word. Therefore, we do not assume these differences to influence the responses to the consonant clusters. Note also that we included the phonetic parameters in the previous analysis by Ulbrich et al. (2016). In this analysis, phonetic parameters did not influence the over-all results and their interpretation.

4.4 Procedure

The nonce words were introduced as names for physical objects, presented as pictures of rare and unfamiliar items such as artifacts, rare animals, insects, reptiles and plants. These were drawn from a number of different websites to ensure that it was unlikely for participants to know an actual name for them. Picture size was standardized (425 x 425 pixels, 15 x 15 cm) for the display on a screen.

The experiment consisted of two parts; a stimulus-presentation phase and a response-elicitation phase. In the first phase, experimental stimuli were auditorily presented via loudspeakers as names for 252 objects which were presented visually on a computer screen to the participants.

Prior to the experiment, participants had to complete a questionnaire to obtain sociodemographic data, they were given instructions regarding the procedure and when necessary feedback and they had to complete a training sequence with 21 practice trials (the equivalent to one block of the actual experiment, see below). The training was repeated when subjects or experimenter felt the need for repetition. Each subject was provided with a different version of the experiment. Stimuli were randomised to exclude effects of order or fatigue. In half of the blocks correct and incorrect parings of words and pictures was reversed, and so was the assignment of correct and incorrect responses to the left joystick button to avoid a handedness bias. In other words, in 12 duplicated versions we assigned the correct response button to the right joystick button. This procedure led to 22 different versions of the experiment.

Participants were comfortably seated in front of a computer screen, in the sound-attenuated cabin. The experiment took approximately 60 minutes including training sequence and self-paced breaks between the experimental blocks. In preparation for the presentation 252 word-picture pairs were quasi-randomized and divided into 12 blocks of 21 trails, each block containing a comparable number of items per condition. During the experiment initially, 21 word-picture pairs were presented, i.e. one block at the time, and the participants had to memorize as many of the word-picture pairs as possible. Following this presentation, the same set of 21 word-picture pairs was presented again, however, only half of the word-picture pairs matched the presentation previously presented in the first phase. The other half of them was matched incorrectly. For each stimulus in the elicitation phase, the participants had to decide as quickly as possible whether the word-picture pair corresponded to that previously introduced in the presentation-phase by pressing the respective joystick buttons. The learning of a new name for an unusual object constitutes an ecologically valid verbal task which ensured that participants did not focus explicitly on the phonotactic properties of the auditory stimuli. The auditory and the visual stimuli were presented simultaneously. Length of stimuli varied (see Section 4.3) and the picture remained on the screen for 1500 ms. The response-interval was opened at the vocalic nucleus of the individual nonce words, since characteristics of the final consonant clusters may be processed as early as during the perception of the vocalic nucleus. The interval had a time-out of 2000 ms. Pressing the joystick buttons triggered the next trial after 1500 ms. Each phase, the presentation-phase and the elicitation-phase, was initiated by an auditory presentation of a synthesised sine wave at 340 Hz for 500 ms.

5 Hypotheses

The design of our experiment and stimuli allows for a number of predictions with respect to the processing of the factors of formedness and existence of the consonant clusters (see Section 2). Differences between individual and crossed conditions are expected to become apparent in both processing speed and the accuracy of recollection of word-picture pairs. In addition, the language-specific influence of the L2 German is expected to differ according to the phonotactic regularities of the respective L1 Russian and Chinese. Finally, we expect an increase of the L2 influence depending on language mastery to be observed in the comparison between the two groups of Chinese learners.

In the comparison of the Russian and the Chinese groups of advanced German learners, we expect sonority to play a less significant role in the processing speed and the responses accuracy of L1 Russian speakers compared to the L1 Chinese speakers. In the comparison between the two groups of L1 Chinese speakers, we expect differences in the processing time and the accuracy of words with well-formed clusters compared to those with ill-formed clusters. In the two groups of advanced speakers we expect nonce words with existent clusters to be processed faster and more often recollected correctly. Predictions are summarized for the three groups in Tables 3, 4a/b.

Tab. 3: Correctness rates and reaction time in recollection of word-picture pairs; L1 Russian

well-formed	=	ill-formed
existent	>	non-existent

Tab. 4: Correctness rates and reaction time in recollection of word-picture pairs; L1 Chinese, advanced and beginning learners

a. advanced learners		
well-formed	>	ill-formed
existent	>	non-existent
b. beginning learners		
well-formed	>	ill-formed
existent	=	non-existent

6 Results

On the basis of reaction times and accuracy in a recollection task of word-picture pairs, the aim of the present study was to test the interplay between phonotactic regularities and input factors in second language learning. We present the behavioural results followed by possible explanations of observed differences in different groups of L2 German learners in the subsequent sections.

Statistical analysis was performed in the R environment for statistical computing (R Core Team 2012) using the *glm2* package for recollection accuracy and the *lme4* package for reaction time (e.g., Baayen 2008; Baayen, Davidson, and Bates 2008). The analysis was carried out in order to test whether phonotactic regularities of an L1 act as a phonological filter for the processing of L2 consonant clusters or if mere expose suffices in order to allow for L2 learning. Furthermore, we were interested if language mastery with the L2 clusters dynamically shapes processing mechanisms in L2 learning. In order to answer these questions, two separate analyses were performed. We first ran an analysis with advanced learners of German (Russian and Chinese L1) before we ran an analysis for two groups of Chinese speakers, advanced learners and beginners for German. In order to understand the contribution of the individual factors in the prediction of the variability in the data we fitted models by stepwise removal of all predictors that did not improve the model. All models were subjected to an ANOVA and compared with the Akaike information criterion (AIC): the model with the lowest AIC was chosen. This was only relevant for the analysis presented in Section 6.4 below, since only in the comparison of reaction times between Chinese advanced learners of L2 German and beginners the predictor Existence (see below) did not improve the model significantly. The regression results were subjected to pairwise comparisons using the *difflsmeans* functions in the *lmerTest* package (Kuznetsova, Brockhoff, and Christensen 2016). The results are visualized with the package *visreg* (Breheny and Burchett 2016). In order to answer the research questions, two separate analyses were performed. We first ran an analysis with advanced learners of German (Russian and Chinese L1) before we ran an analysis for two groups of Chinese speakers, advanced learners and beginners for German.

6.1 L1 Influence: accuracy

Accuracy of recollection as dependent binomial variable was analyzed in a general linear model (Dobson 2010). The model was fitted with the predictors L1 (advanced Russian and Chinese learners), FORMEDNESS (WF and IF clusters) and EXISTENCE (EX and NX clusters); SUBJECT and ITEM were included as random factors. Results for the test of accurate recollection of word-picture pairs are presented in Table 5 and illustrated in Figure 1. The analysis revealed that main effects were found for FORMEDNESS and L1. Significant interactions were found for L1 and EXISTENCE and for L1 and FORMEDNESS. Also, a significant three-way interaction was found for L1, EXISTENCE and FORMEDNESS (see Table 5).[4]

4 In accordance with recommendations by Baayen et al. (2008), we take t-values t ≥ |2| to indicate a significant result.

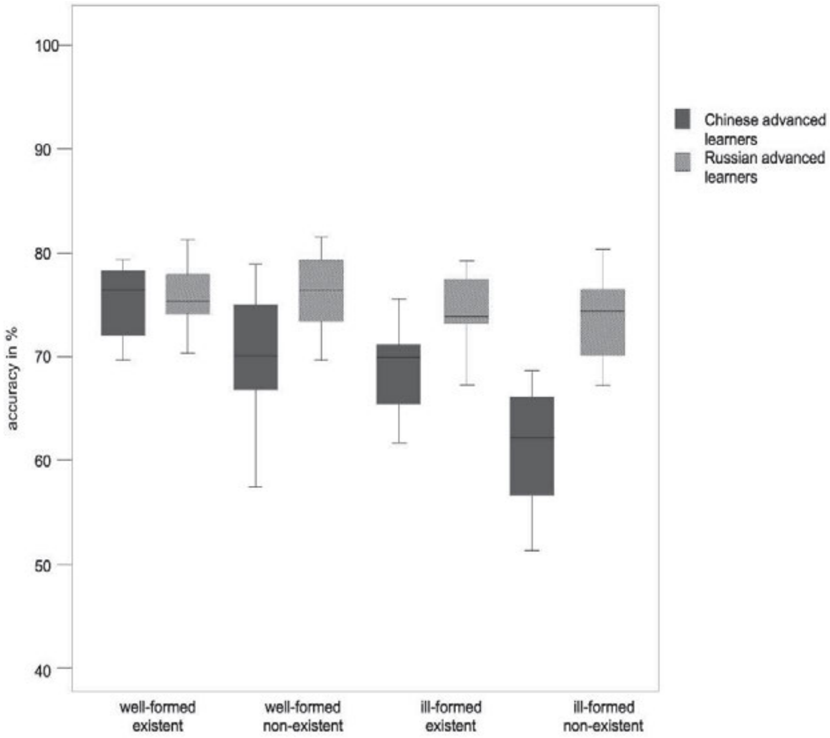


Fig. 1: Recollection accuracy (%) obtained for Russian and Chinese advanced learners

Tab. 5: Main effects and interactions for accuracy, Russian and Chinese advanced learners

	Estimate	SE	z-value	Pr(>|z|)
(Intercept) Russian, WF, EX	2.97681	0.18727	7.532	< 2e−16 ***
Chinese	0.53242	0.26352	2.122	0.0032123 **
IF	0.3462	0.2034	2.998	0.024231 *
NX	0.3546	0.2227	2.432	0.019867 *
Chinese : IF	0.35421	0.2997	2.876	0.002143 **
Chinese : NX	0.46726	0.3672	2.184	0.018988 *
IF : NX	0.13244	0.2176	1.989	0.045234 *
Chinese : IF : NX	0.78447	0.4222	2.872	0.001231 **
Significance codes: 0.0001 '***' 0.001 '**' 0.01 '*'				

Overall, Russian learners of German were better in the recollection of word-picture pairs compared to Chinese learners. Post-hoc test revealed that, contrary to the Chinese advanced learners of German, EXISTENCE and FORMEDNESS did not significantly influence the performance of the Russian participants (see Table 6). The interaction of L1 and FORMEDNESS results from the fact that Chinese learners more accurately recollected word-picture pairs with words containing WF clusters. The factor EXISTENCE was only significant for IF clusters of Chinese advanced learners. The three-way interaction of L1, FORMEDNESS and EXISTENCE can be attributed to the fact that Chinese and Russian participants' performance in the recollection task did not differ in the recollection of words with WF-EX clusters.

Tab. 6a: Pairwise comparison of accuracy, factors FORMEDNESS and EXISTENCE in Russian learners of L2 German

	Estimate	SE	z-value	Pr(>|z|)
Russian				
WF-NX – WF-EX	0.3438	0.06381	0.9142	0.3223
WF-NX – IF-EX	−0.2176	0.03119	−0.8882	0.2678
WF-NX – IF-NX	−0.3451	0.03763	−1.9352	0.2778
IF-EX – IF-NX	0.1133	0.10091	0.9111	0.8856
IF-EX – WF-EX	−0.4457	0.01132	−1.2273	0.3223
IF-NX – WF-EX	0.1568	0.01232	0.9646	1.0000

Tab. 6b: Pairwise comparison of accuracy, factors FORMEDNESS and EXISTENCE in Chinese learners of L2 German

	Estimate	SE	z-value	Pr(>|z|)
Chinese				
WF-NX – WF-EX	−0.1521	0.05228	−1.9813	0.2112
WF-NX – IF-EX	0.1678	0.09872	0.9823	0.7784
WF-NX – IF-NX	0.7382	0.19233	5.1436	<0.01 *
IF-EX – IF-NX	0.3152	0.08928	5.9871	<0.01*
IF-EX – WF-EX	−0.3193	0.09328	−7.3427	<0.01 *
IF-NX – WF-EX	−0.4554	0.18311	−1.6323	1.0000
Significance code: 0.01 '*'				

6.2 L1 Influence: Reaction time

For the statistical analysis of reaction times we used the *lme4* package in the R environment to calculate a generalized linear mixed-effects model (Bates 2005; Bates and Sarkar 2006). The model was again fitted with L1 (advanced Russian and Chinese learners), FORMEDNESS (WF and IF clusters) and EXISTENCE (EX and NX clusters) as fixed factors; SUBJECT and ITEM were included as random factors. T-values > 2 or < –2 indicate significant results (Baayen, Davidson, and Bates 2008). Reaction time data are presented in Figure 2. The analysis of these data (Table 7) reveals significant main effects for FORMEDNESS, L1 and EXISTENCE, and significant interactions between all three factors.

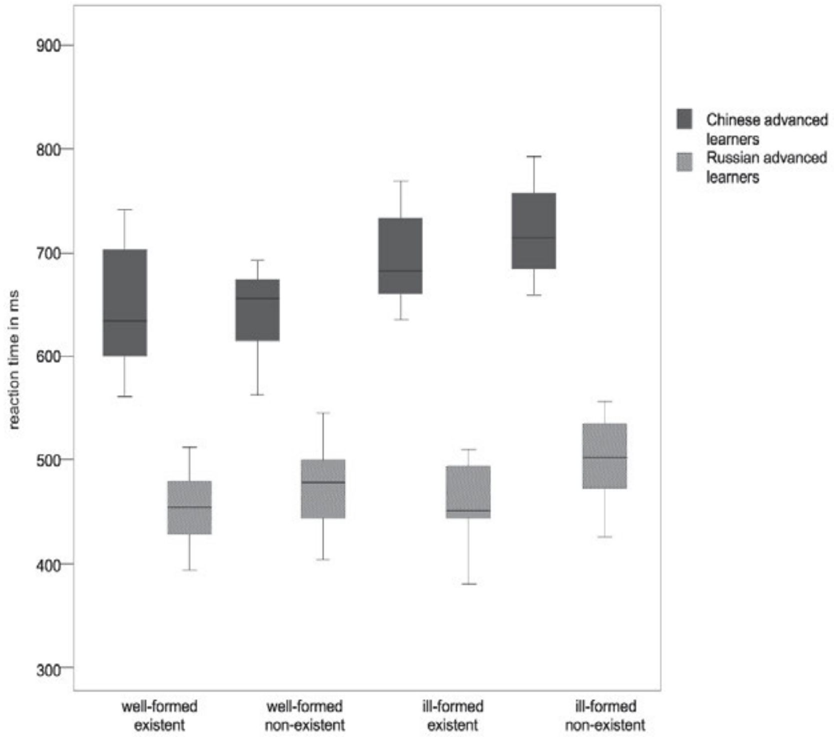


Fig. 2: Reaction times (ms) obtained for Russian and Chinese advanced learners

Tab. 7: Main effects and interactions for reaction times; Russian and Chinese advanced learners

	Estimate	SE	t-value	Pr(>|t|)
(Intercept) Russian, WF, EX	**456.783**	**22.917**	**23.635**	**< 2e–16 *** **
Chinese	**189.736**	**42.2143**	**2.902**	**0.000354 *** **
IF	**23.776**	**28.6527**	**2.035**	**0.043334 ***
NX	**24.652**	**17.9862**	**2.156**	**0.012577 ***
Chinese : IF	**42.554**	**27.8787**	**1.952**	**0.048722 ***
Chinese : NX	13.421	19.1124	1.765	0.054121
IF : NX	**32.625**	**21.6571**	**2.012**	**0.038952 ***
Chinese : IF : NX	**34.927**	**30.003**	**2.542**	**0.024521 ***

Significance codes: 0.0001 '***' 0.01 '*'

Overall, Russian learners of German performed the recollection task faster compared to the proficiency-matched group of Chinese learners. A significant two-way interaction was found for L1 and FORMEDNESS; no difference between words with WF clusters was found in the response times obtained in the Russian group of participants, whereas recollection was faster in the response time to stimuli with WF clusters in the Chinese participants' group. Furthermore, a significant interaction between FORMEDNESS and EXISTENCE can be explained by the fact that responses to stimuli with IF-NX clusters are significantly slower than to those with WF-EX and IF-EX, as revealed by the pairwise comparisons in Table 8. In contrast, Chinese speakers' responses to stimuli with IF-EX and IF-NX clusters did not differ significantly, explaining the significant three-way interaction between FORMEDNESS, EXISTENCE and L1.

Tab. 8a: Pairwise comparisons of reaction time, factors FORMEDNESS and EXISTENCE in Russian learners of L2 German

	Estimate	SE	z-value	Pr(>|z|)
Russian				
WF-NX – WF-EX	13.216	9.821	1.265	0.5313
WF-NX – IF-EX	−16.299	12.638	−2.871	<0.01 *
WF-NX – IF-NX	21.351	15.481	1.832	0.2322
IF-EX – IF-NX	23.866	15.882	2.034	<0.01 *
IF-EX – WF-EX	−8.993	11.163	−1.356	0.1987
IF-NX – WF-EX	18.083	16.726	1.812	0.5112
Significance codes: 0.01 '*'				

Tab. 8b: Pairwise comparison of reaction time, factors FORMEDNESS and EXISTENCE in Chinese learners of L2 German

	Estimate	SE	z-value	Pr(>|z|)
Chinese				
WF-NX – WF-EX	9.441	11.452	1.176	0.1011
WF-NX – IF-EX	16.443	15.847	2.778	< 0.01 *
WF-NX – IF-NX	25.892	16.837	4.231	< 0.01 *
IF-EX – IF-NX	18.233	12.582	1.231	0.7221
IF-EX – WF-EX	23.232	11.991	3.002	< 0.01 *
IF-NX – WF-EX	29.998	12.643	2.192	< 0.01 *
Significance codes: 0.01 '*'				

6.3 L2 Proficiency: Recollection

In a second analysis, we compared different levels of L2 proficiency applying the same statistical analysis as used in the cross-linguistic comparison of Russian and Chinese advanced learners to the two groups of L1 Chinese learners of L2 German, i.e., beginners and advanced learners. We calculated a generalized linear model for accuracy and a generalized mixed effects model for reaction times. As the issue of the proficiency was addressed only in the data of Chinese learners of L2 German, we ran the analyses with the predictor L2-PROFICIENCY (beginners and advanced learners). As in the previous analysis, we were interested in the contribution of the predictors FORMEDNESS and EXISTENCE, so we fit the model with FORMEDNESS and EXISTENCE as fixed factors. SUBJECT and ITEM were again included as random factors. For the analysis of the interaction, we ran pairwise comparisons for the subsets of advanced learners and beginners.

The results for the comparison of the two Chinese groups of L2 learners differing in their level of language mastery are presented in Figure 3. The dependent binomial variable was the accuracy of responses. The model included the binary fixed factors of L2-PROFICIENCY, FORMEDNESS and EXISTENCE. SUBJECT and ITEM were included as random factors. Main effects were found for FORMEDNESS, EXISTENCE and L2-PROFICIENCY, and significant two-way interactions for L2-PROFICIENCY and EXISTENCE, FORMEDNESS and EXISTENCE. The three factors are also involved in a three-way interaction (see Table 9).

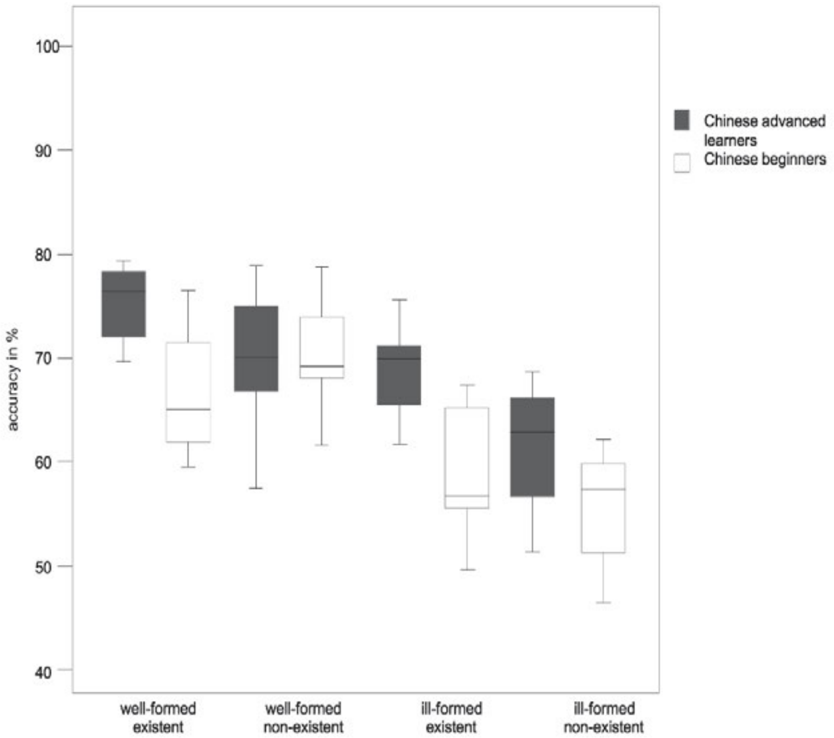


Fig. 3: Recollection accuracy (%) obtained for L1 Chinese beginners and advanced learners

Tab. 9: Main effects and interactions for accuracy, Chinese beginners and advanced learners

	Estimate	**SE**	**z-value**	**Pr(>|z|)**
(Intercept) Beginners, WF, EX	1.824	0.278	9.347	<2e–16 ***
Advanced learners	**0.348**	**0.19811**	**1.982**	**0.04901 ***
IF	**0.562**	**0.30363**	**2.036**	**0.02456 ***
NX	**0.125**	**0.21871**	**1.892**	**0.19652 ***
Advanced learners : IF	0.243	0.30221	1.072	0.08731
Advanced learners: NX	**0.663**	**0.54219**	**2.962**	**0.04552 ***
IF : **NX**	**0.725**	**0.57339**	**2.342**	**0.02571 ***
Advanced learners : IF : NX	**0.872**	**0.72653**	**2.972**	**0.03752 ***
Significance codes: 0.0001 '***' 0.01 '*'				

Overall, the group of advanced learners performed better in the recollection task of word-picture pairs. In addition, the recollection was more successful in nonce words with WF clusters compared to those with IF clusters. The main effect for Existence is only barely significant but this factor features in all of the observed interactions. The significant interaction between L2-Proficiency, Formedness and Existence is due to the fact that, unlike the beginners, advanced learners perform better in the recollection of word-picture pairs containing nonce words with EX clusters. Beginners are even better in the recollection of nonce words with NX clusters but only if the clusters are well-formed. Note that there is no difference between the two groups in the performance of the task when WF-NX clusters are concerned. Finally, nonce words with WF-NX clusters and their counterpart, IF-EX clusters show no difference in the advanced learners group (Table 10).

Tab. 10a: Pairwise comparison of accuracy, factors FORMEDNESS and EXISTENCE in advanced Chinese learners of L2 German

	Estimate	SE	z-value	Pr(>|z|)
Advanced				
WF-NX – WF-EX	−0.938331	0.097463	−6.832	<0.01 *
WF-NX – IF-EX	0.562683	0.082741	3.722	0.261
WF-NX – IF-NX	−0.927483	0.088372	3.988	<0.01 *
IF-EX – IF-NX	−0.798434	0.099726	−5.973	<0.01 *
IF-EX – WF-EX	−0.922572	0.092216	−7.839	<0.01 *
IF-NX – WF-EX	−0.737249	0.086113	−6.322	<0.01 *
Significance code: 0.01 ‘*’				

Tab. 10b: Pairwise comparison of accuracy, factors FORMEDNESS and EXISTENCE in Chinese beginners of L2 German

	Estimate	SE	z-value	Pr(>|z|)
Beginners				
WF-NX – WF-EX	0.722634	0.183726	−2.231	<0.01 *
WF-NX – IF-EX	−0.726148	0.098262	6.082	<0.01 *
WF-NX – IF-NX	−0.973265	0.073672	5.916	<0.001 **
IF-EX – IF-NX	−0.187634	0.887247	1.123	0.3632
IF-EX – WF-EX	−0.814212	0.086727	−7.813	<0.01 *
IF-NX – WF-EX	−0.991338	0.078782	−3.928	<0.01 *
Significance codes: 0.001 ‘**’, 0.01 ‘*’				

6.4 L2 Proficiency: Reaction times

Again, we used the *lme4* package in the R environment to calculate a generalized linear mixed-effects model (Bates 2005; Bates and Sarkar 2006) for the statistical analysis of reaction times within the data set of Chinese learners. In this analysis we fitted the model with L2-PROFICIENCY (advanced learners and beginners), FORMEDNESS (WF and IF clusters) and EXISTENCE (EX and NX clusters) as fixed factors and SUBJECT and ITEM as random factors. T-values > 2 or < −2 indicate significant results (Baayen, Davidson, and Bates 2008). The analysis revealed significant main effects for L2-PROFICIENCY and FORMEDNESS only. No significant main effect was found for EXISTENCE, and the factor was not involved in a significant interaction. We therefore fitted a model with the predictors of FORMEDNESS and L2-PROFICIENCY only and subsequently compared the two models in an ANOVA. The model without EXISTENCE as predictor had the lower AIC. The data are presented in Figure 4.

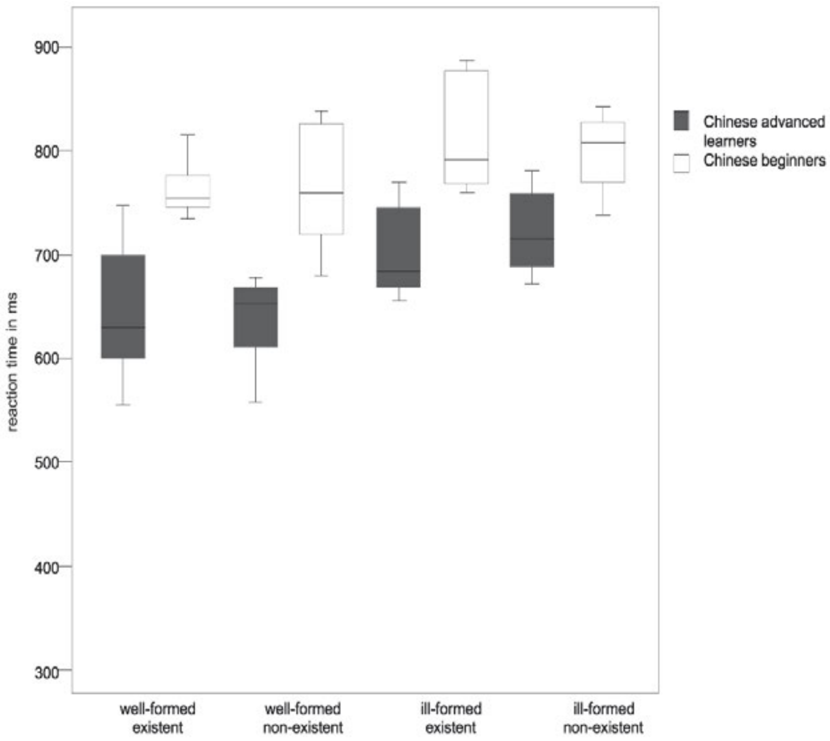


Fig. 4: Reaction times (ms) obtained for L1 Chinese beginners and advanced learners

Not surprisingly, L1 Chinese advanced learners of German performed the task faster (see Table 11). Both groups of learners were faster in the recollection of word-picture pairs containing WF clusters compared to those with IF clusters.

Tab. 11: Main effects and interactions for reaction times, Chinese beginners and advanced learners

	Estimate	SE	t-value	Pr(>|t|)
(Intercept) Beginners, WF	764.826	39.52	38.218	<2e–19 ***
Advanced learners	**–64.226**	**22.552**	**–2.021**	**0.00121 ** **
IF	**39.974**	**19.732**	**2.004**	**0.08371 * **
Advanced learners : IF	**–17.159**	**.831**	**–0.927**	**0.01643* **

Significance codes: 0.0001 '***' 0.001 '**' 0.01 '*'

7 Discussion

The aim of the present study was to investigate the relative role of phonotactic principles and input factors in the processing of final consonant clusters in L2. More specifically, we were interested whether either sonority or existence show stronger impact on the success of L2 mastery, or if in fact both factors play a role. By using two groups of L2 learners with L1s that not only vary considerably in their sonority requirements from the target language German, but also from each other, we approached the question how processing of clusters is affected by the phonotactic complexity of final clusters in the L1. Finally, to figure out whether or not the proficiency level affects the processing of phonotactic requirements, we obtained data from L1 Chinese beginners and advanced learners of German.

The results reveal several significant differences between the two groups of advanced learners of L2 German. Russian participants' processing time was faster regardless of the adherence or violation to the sonority requirements of German. The relative frequent occurrence of violations of the sonority requirements in the L1 is transferred into the processing of L2, still at a relative high level of L2 proficiency. However, the higher rate of correct recollection and the faster processing speed of word-picture pairs involving words with well-formed clusters suggest that Russian speakers are sensitive to sonority violations, in spite of the fact that they have frequently encountered ill-formed clusters in their L1. A possible explanation is, in line with previous accounts, that language-specific phonotactic requirements of an L2 not represented in the L1 can be internalized by proficient L2 learners (Mikhaylova 2009; Schwartz and Sprouse 1996). This means that with increasing language mastery the influence of an L1 phonological filter is reduced. It may in fact not apply at all in case it is more flexible than that of the L2 and therefore cannot restrain the phonotactics of the L2. More specifically, our results suggest that the impact of L2 phonotactic constraints depends on the relative "restrictive power" of the two competing phonological filters of L1 and L2. If the phonological filter of the L1 constrains phonotactic complexity more compared to the L2 filter, as in the case of L1 Russian and L2 German, the L2 filter's impact is more limited than in the case of a less constraining phonological L2 filter, as in the case of L1 Chinese and L2 German. In order to investigate whether the impact of the L2 filter affects the performance level rather than the competence level, and in order to gain a better understanding of the dynamics between the L1 and the L2 filter, longitudinal data would be best suited in addition to very detailed proficiency testing.

Another explanation for the sensitivity to sonority as observed in the Russian and the Chinese participant groups is that the sonority hierarchy actually has universal applicability, as argued by Berent et al. (2010, 2014). This view is supported by two comparable EEG studies by Ulbrich et al. (2016) and Wiese et al. (2017) demonstrating a dominant role for sonority over existence in the processing of German and Polish coda consonant clusters.

However, we also found a significant influence of existence in the processing speed and the accuracy in response to the consonant clusters. The influence was not consistent though. In fact, both Russian and Chinese advanced learners responded faster, and the beginner group of Chinese L2 German learners responded more accurately to stimuli with well-formed and non-existent clusters. A similar observation has previously been attributed to the assumption that well-formed clusters are expected to exist whereas ill-formed clusters are more likely to not exist. Due to such contradicting information conflicts may occur and slow down processing transitions from pre-lexical processing to lexical access as suggested in previous work by Ulbrich et al. (2016) and Wiese et al. (2017).

Chinese advanced learners show high sensitivity to the sonority requirements in both of the measures, i.e., the reaction time and the recollection rate. Responses to nonce words with well-formed clusters were faster and more frequently accurate than to those with ill-formed clusters.[5] Chinese does not allow for more than one consonant in syllable-final position, so that the sensitivity to the sonority requirement apparent in German final consonant clusters could only be due to the experience with input patterns from the L2. However, similar results were obtained for the Chinese groups of beginners. This illustrates the role of the

5 Note that this effect cannot be overridden by input frequency since some of the existing ill-formed clusters had to be presented six times and no difference was found in either processing speed or correctness rate between them and those presented only three times.

sonority requirement as a universal phonotactic constraint: both groups are sensitive to the sonority principle of final CC clusters, even though their L1 does not allow CC clusters in coda position. Exposure to actually occurring clusters still has an influence, since advanced Chinese learners of L2 German process word-picture pairs with non-existent clusters with significantly lower correctness rates than those with existent clusters. In the reaction time and the correctness rate for Chinese beginners, existence does not play a significant role. This interaction, again, provides evidence for a role of both the sonority principle and the exposure. Overall, the results indicate that L2 learners benefit from both, universal principles and input frequency, in the process of the acquisition of phonotactic information in their L2 (see also Archibald 2004; Dresher 1999).

8 Conclusions

Consonantal clusters have been argued to provide crucial information on the beginnings and/or endings of words (Trubetzkoy 1967; Weber and Cutler 2006). Thus, their speedy processing is immensely helpful for one central task in speech comprehension, namely the segmentation of continuous speech into word-like units, and any hypothesis derived from the processing of such clusters is useful. The aim of the study was to examine the influence of the phonotactic constraint of sonority as well as the influence of the existence of specific consonant clusters by means of reaction time measures and correctness rates in a word-picture matching task. The present results show that both, formedness and existence, play a significant role in the processing of word-like items in L2 German by learners with considerably different phonotactic requirements in their native language. In accordance with previous findings, we demonstrated that implicit knowledge of phonological universals as well as frequent exposure influences the online processing of structures. In addition, L2 constraints may even shape language-specific phonotactic mechanisms in both directions, i.e. expand language-specific restrictions, as in the case of L1 Chinese speakers, or limit them, as in the case of Russian L2 learners of German.

Our results therefore provide evidence for both views argued for in current phonological theory (and often seen as mutually exclusive): phonological knowledge is based both on abstract principles such as sonority and on input patterns such as prior existence. Identifying the correct weighting for the influences of UG and input factors remains, of course, an empirical question, to be determined by further future study.

Acknowledgements: The present study was supported by a grant from the Hessian Ministry of Research and the Arts within the LOEWE program *Fundierung linguistischer Basiskategorien*. We thank Ingo Plag and Antonia Götz for their valuable advice.

References

Archibald, John. 2004. Interfaces in the prosodic hierarchy: New structures and the phonological parser. International Journal of Bilingualism 8(1), 29–50.
Baayen, Harald. 2008. Analyzing Linguistic Data: A Practical Introduction to Statistics Using R. Cambridge: Cambridge University Press.
Baayen, Harald, Douglas J. Davidson & Douglas M. Bates. 2008. Mixed-Effects Modeling with Crossed Random Effects for Subjects and Items. Journal of Memory and Language 59. 390–412.
Bates, Douglas. 2005. Fitting linear mixed models in R. R News 5. 27–30.
Bates, Douglas & Deepayan Sarkar. 2006. lme4: Linear Mixed-Effects Models Using S4 Classes. Retrieved from https://cran.r-project.org, last accessed 25.07.2017.
Benson, Cathy. 2002. Transfer/cross-linguistic influence. ELT Journal 56(1). 68–70.
Berent, Iris, Evan Balaban, Tracy Lennertz & Vered Vaknin-Nusbaum. 2010. Phonological universals constrain the processing of nonspeech stimuli. Journal of Experimental Psychology: General 139(3). 418–435.
Berent, Iris & Tracy Lennertz. 2010. Universal Constraints on the Sound Structure of Language: Phonological or Acoustic? Journal of Experimental Psychology: Human Perception and Performance 36. 212–223.
Berent, Iris, Hong Pan, Xu Zhao, Jane Epstein, Monica L. Bennett, Vibhas Deshpande, Ravi T. Seethamraju & Emily Stern. 2014. Language Universals Engage Broca's Area. PloS ONE 9(4). e95155.
Bley-Vroman, Robert. 1989. The Logical Problem of Foreign Language Learning. In Susan Gass & Jacquelyn Schachter (eds.), Linguistic Perspectives on Second Language Acquisition, 41–68. Cambridge: Cambridge University Press.
Boll-Avetisyan, Natalie. 2018. The role of phonological structure in speech segmentation by infants and adults: a review and methodological considerations. In Christiane Ulbrich, Alexander Werth & Richard Wiese (eds.), Empirical approaches to the phonological structure of words (Linguistische Arbeiten 567), 213–240. Berlin & Boston: De Gruyter.
Breheny, Patrick & Woodrow Burchett. 2016. Visualization of regression models using visreg. Version 2.3-0. http://myweb.uiowa.edu/pbreheny/publications/visreg.pdf (accessed 25 July 2017).
Brown, Cynthia A. 1998. The role of the L1 grammar in the L2 acquisition of segmental structure. Second Language Research 14. 136–193.
Bybee, Joan & Paul J. Hopper. 2001. Frequency and the emergence of linguistic structure. Amsterdam: John Benjamins.
Chang, Charles B. 2010. First language phonetic drift during second language acquisition (Ph.D. Dissertation). Berkeley: University of California.
Cook, Vivian. 1991. The poverty-of-the-stimulus argument and multi-competence. Second Language Research 7(2). 103–117.

Cook, Vivian. 2003. Effects of the second language on the first. Clevedon, UK: Multilingual Matters.
Davidson, Lisa & Kevin Roon. 2008. Durational correlates for differentiating consonant sequences in Russian. Journal of the International Phonetic Association. 38. 137–165.
De Bot, Kees. 1986. The transfer of intonation and the missing database. In Eric Kellerman & Michael S. Smith (eds.), Crosslinguistic Influence in Second Language Acquisition, 110–119. New York: Pergamon.
Dobson, A. J. 2010. An Introduction to Generalized Linear Models, 2nd edn. New York: Chapman & Hall/CRC Press.
Domahs, Ulrike, Wolfgang Kehrein, Johannes Knaus, Richard Wiese & Matthias Schlesewsky. 2009. Event-related Potentials Reflecting the Processing of Phonological Constraint Violations. Language and Speech 52(4), 415–435.
Dresher, Bezalel E. 1999. Charting the learning path: Cues to parameter setting. Linguistic Inquiry 30. 27–67.
Duanmu, San. 2002. The Phonology of Standard Chinese. Oxford: Oxford University Press.
Duanmu, San. 2014. Syllable structure and stress. In C. T. James Huang, Y. H. Audrey Li & Andrew Simpson (eds.), The handbook of Chinese linguistics, 422–442. Hoboken, NJ: John Wiley & Sons.
Ellis, Nick C. 2005. At the interface: Dynamic interactions of explicit and implicit language knowledge. Studies in Second Language Acquisition 27. 305–352.
Ellis, Nick C. & Diane Larsen-Freeman. 2006. Language emergence: Implications for applied linguistics—Introduction to the special issue. Applied Linguistics 27(4). 558–589.
Ellis, Rod. 1994. The Study of Second Language Acquisition. Oxford: Oxford University Press.
Flege, James E. 2002. Interactions between the native and second-language phonetic systems. In Petra Burmeister, Thorsten Piske & Andreas Rohde (eds.), An integrated view of language development: Papers in honor of Henning Wode, 217–244. Trier: Wissenschaftlicher Verlag Trier.
Grosjean, François. 1998. Transfer and language mode: Commentary on Natascha Müller "Transfer in bilingual first language acquisition". Bilingualism: Language and Cognition 1. 175–176.
Halicki, Shannon. 2010. L2 Knowledge of Target Phonotactics: Learner Judgments of French Resuffixation. In Matthew T. Prior, Yukiko Watanabe & Sang-Ki Lee (eds.), Selected Proceedings of the 2008 Second Language Research Forum, 166–180. Somerville, MA: Cascadilla Proceedings Project.
Hall, Tracy A. 1992. Syllable Structure and Syllable Related Processes in German. Tübingen: Niemeyer.
Halle, Pierre A., Juan Segui, Uli Frauenfelder & Christine Meunier. 1998. Processing of illegal consonant clusters: A case of perceptual assimilation? Journal of Experimental Psychology: Human Perception and Performance 24. 592–608.
Heuven, Walter J. B. van, Ton Dijkstra & Jonathan Grainger. 1998. Orthographic neighborhood effects in bilingual word recognition. Journal of Memory and Language 39. 458–483.
Jared, Debra & Judith Kroll. 2001. Do bilinguals activate phonological representations in one or both of their languages when naming words? Journal of Memory and Language 44. 2–31.
Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen Christensen. 2016. lmerTest: tests in Linear mixed effects models. R package version 2.0–33.
Major, Roy C. 1994. Chronological and Stylistic Aspects of Second Language Acquisition of Consonant Clusters. Language Learning 44(4). 655–680.

Mennen, Ineke. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek. Journal of Phonetics 32(4). 543–563.
Mikhaylova, Anna. 2009. L2 Nonword Recognition and Phonotactic Constraints. University of Pennsylvania Working Papers in Linguistics 15(1). Issue 1: Proceedings of the 32nd Annual Penn Linguistic Colloquium. Article 17.
Moisik, Scott. 2009. Frequency, phonology, the brain, and second language acquisition: Laying the ground work for a hybrid approach. Working Papers of the Linguistics Circle. 164–188.
Moreton, Elliott. 2002. Structural constraints in the perception of English stop-sonorant clusters. Cognition 84. 55–71.
Moreton, Elliott & Joe Pater. 2012. Structure and substance in artificial-phonology learning. Part I, Structure. Part II, Substance. Language and Linguistics Compass 6(11). 686–701 and 702–718.
Myers, Scott & Jaye Padgett. 2014. Domain generalisation in artificial language learning. Phonology. 31(3). 399–433.
Nooteboom, Sieb. 1997. The prosody of speech: melody and rhythm. The handbook of phonetic sciences. 5. 640–673.
Odlin, Terence. 1989. Language Transfer: Cross-linguistic influence in language learning. Cambridge: Cambridge University Press.
Parker, Steve. 2012. The Sonority Controversy. Berlin & Boston: Walter de Gruyter.
Pavlenko, Aneta. 2000. L2 influence on L1 in late bilingualism. Issues in Applied Linguistics 11(2). 175–205.
Pavlenko, Aneta & Scott Jarvis. 2002. Bidirectional transfer. Applied Linguistics 23(2). 190–214.
R Core Team. 2012. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/ (accessed 25 July 2017).
Rietveld, Antonius C. M. & Carlos Gussenhoven. 1985. On the Relation between Pitch Excursion Size and Prominence. Journal of Phonetics 13. 299–308.
Schwartz, Bonnie & Rex A. Sprouse. 1996. L2 cognitive states and the Full Transfer/Full Access model. Second Language Research 12. 40–72.
Sievers, Eduard. 1901. Grundzüge der Phonetik zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen. Leipzig: Breitkopf & Härtel.
Strange, Winifred. 1995. Cross-language study of speech perception: a historical review. In Winifred Strange (ed.), Speech Perception and Linguistic Experience: Issues in Cross-Language Research, 3–45. Timonium, MD: York Press.
Švedova, Natalija J. 1980. Russkaja grammatika. Tom I. Fonetika, fonologija, udarenie, intonacija, slovoobrazzovanije, morfologija. [Russian Grammar, Volume 1. Phonetic, Phonology, Stress, Intonation, Word Formation, Morphology]. Moskva: Nauka.
‘t Hart, Johan. 1981. Differential sensivity to pitch distance, particularly in speech. Journal of Acoustical Society of America 69(3). 811–821.
‘t Hart, Johan, René Collier & Antonie Cohen. 1990. A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody. Cambridge: Cambridge University Press.
Timberlake, Alan. 2004. A Reference Grammar of Russian. Cambridge: Cambridge University Press.
Trubetzkoy, Nikolai S. 1939/1967. Grundzüge der Phonologie, 2nd edn. Göttingen: Vandenhoeck & Ruprecht.

Ulbrich, Christiane. 2013. German pitches in English: Production and perception of cross-varietal differences in L2. Bilingualism: Language and Cognition 16(2). 397–419.
Ulbrich, Christiane, Philipp M. Alday, Johannes Knaus, Paula Orzechowska, Richard Wiese. 2016. The role of phonotactic principles in language processing. Language, Cognition and Neuroscience 31(5). 662–682. DOI: 10.1080/23273798.2015.1136427.
Ulbrich, Christiane & Mikhail Ordin. 2014. Can L2-English influence L1-German? The case of post-vocalic /r/. Journal of Phonetics 45. 26–42.
Vennemann, Theo. 1972. On the theory of syllabic phonology. Linguistische Berichte 18(1). 1–18.
Weber, Andrea & Anne Cutler. 2006. First-language phonotactics in second-language listening. Journal of the Acoustic Society of America 119. 597–607.
Whitney, William D. 1865. On the Relation of Vowels and Consonants. Journal of the American Oriental Society 8. 357–373.
Wiese, Richard. 1988. Silbische und Lexikalische Phonologie: Studien zum Chinesischen und Deutschen. Tübingen: Niemeyer.
Wiese, Richard. 1997. Underspecification and the description of Chinese vowels. In Wang Jialing & Norval Smith (eds.), Studies in Chinese Phonology, 219–249. Berlin: Mouton de Gruyter.
Wiese, Richard. 2000. The Phonology of German. Oxford: Oxford University Press.
Wiese, Richard. 2016. Die Rolle der Silbe in der Lautsprache. In Ulrike Domahs & Beatrice Primus (eds.), Handbuch Laut – Gebärde – Buchstabe, 46–63. Berlin: De Gruyter.
Wiese, Richard, Paula Orzechowska, Philipp M. Alday & Christiane Ulbrich. 2017. Structural Principles or Frequency of Use? An ERP Experiment on the Learnability of Consonant Clusters. Frontiers in Psychology 7. 2005.
Wilson, Colin. 2003. Experimental Investigation of Phonological Naturalness. In Gina Garding & Mimu Tsujimura (eds.), Proceedings of the 22nd West Coast Conference on Formal Linguistics, 533–546. Somerville, MA: Cascadilla Press.

Alexander Werth, Marie Josephine Rocholl, Karen Henrich, Manuela Lanwermeyer, Hanni Th. Schnell, Ulrike Domahs, Joachim Herrgen, and Jürgen Erich Schmidt

The interaction of vowel quantity and tonal cues in cognitive processing: An MMN-study concerning dialectal and standard varieties

Abstract: In this study, the influence of two dialectal prosodic features on the processing of lexical meaning during spoken word recognition was investigated in German dialect and non-dialect speakers. Previous studies in the field of German dialectology investigated differences between dialectal varieties and the Standard German variety by using mainly offline production and perception studies. The present study concentrates on brain responses to the phonological contrast of vowel quantity combined with tone accents, which occur in Germany exclusively in the Middle-Franconian dialect area (Moselle-Franconian, Ripuarian and southern Low Franconian dialects) but not in Standard German. In an event-related potential-study using a classic oddball paradigm, two groups of participants (dialect and Standard German speakers) were presented with two

Alexander Werth (Corresponding author): Forschungszentrum Deutscher Sprachatlas, Philipps-Universität Marburg, Pilgrimstein 16, D-35032 Marburg, alexander.werth@uni-marburg.de
Marie Josephine Rocholl: Amazon Development Center, Aureliusstraße 2, D-52064 Aachen, mjroc@amazon.com
Karen Henrich: Max-Planck-Institut für empirische Ästhetik, Grüneburgweg 14, D-60322 Frankfurt am Main, karen.henrich@aesthetics.mpg.de
Manuela Lanwermeyer: Forschungszentrum Deutscher Sprachatlas, Philipps-Universität Marburg, Pilgrimstein 16, D-35032 Marburg, manuela.lanwermeyer@uni-marburg.de
Hanni Th. Schnell: Forschungszentrum Deutscher Sprachatlas, Philipps-Universität Marburg, Pilgrimstein 16, D-35032 Marburg, hanni.schnell@uni-marburg.de
Ulrike Domahs: Philipps-Universität Marburg, Pilgrimstein 16, D-35032 Marburg, ulrike.domahs@uni-marburg.de
Joachim Herrgen: Forschungszentrum Deutscher Sprachatlas, Philipps-Universität Marburg, Pilgrimstein 16, D-35032 Marburg, herrgen@uni-marburg.de
Jürgen Erich Schmidt: Forschungszentrum Deutscher Sprachatlas, Philipps-Universität Marburg, Pilgrimstein 16, D-35032 Marburg, schmidtj@uni-marburg.de

DOI 10.1515/9783110542899-008

words of a minimal pair ([ʃa̠ː^{2}l] 'stale' vs. [ʃa̠l·^{2}] 'acoustic noise') which have inverted lengths for the vowel and the lateral but both bear Tone Accent 2. Late mismatch negativity effects resulting from pre-attentive processing differ in amplitude and latency between the two groups of participants indicating varying phonological relevance of prosodic cues in these two varieties. Although both participant groups perceive Tone Accent 2 as a high tone, only the dialect group uses rules of tone-text-association within the minimal pair for lexical access.

Keywords: vowel quantity, tone accent, dialect, event-related potentials (ERP), auditory evoked responses, late mismatch negativity (MMN), contrast enhancement

1 Introduction

1.1 General Introduction

The Standard German vowel system contains 15 monophthongs and three diphthongs and can therefore be regarded as a relatively rich system of vowel phonemes compared to most other languages.[1] In some German dialects such as Moselle-Franconian, phoneme inventories are even more extensive. In Mayen, which is part of the Moselle-Franconian dialect area, 20 monophthongs and three diphthongs exist in the vowel inventory. While tenseness and long quantity[2] and then laxness and short quantity appear for most of the vowel phonemes in complementary distribution in the Standard German variety, all tense vowels can also appear with short quantity in stressed positions in this dialect (e.g., [ʃtuf] 'living room' and [ʃtof] 'textile fabric'). This means that there are absolutely parallel rows of short and long vowels concerning quality, and therefore vowel quantity has a larger phonological load than in the Standard German variety. Thus, an examination of quantity seems very worthwhile for this particular dialect variety.

In addition, in the Middle-Franconian dialects including Moselle-Franconian, there are tone accents associated with vowels or other sonorants, which can be used for lexical access. But the question still remains how these phonological

1 Some authors such as Wiese (2011) also consider the two vowels [ə] und [ɐ] to be phonemes of Standard German, both of which can only occur in an unstressed position. Also the vowel [ɛː] is discussed concerning its phoneme status. It is considered here, but it is not used by all German speakers and is distinctive only in some minimal pairs.

2 The term quantity is used here only for phonological distinction, whereas length refers to the units of perception and duration to the acoustic units.

differences between the different German varieties influence vowel processing. This question will be examined by means of electrophysiological responses to vowel perception.

Research findings already exist for neural processing of prosodic cues like length (e.g., Näätänen, Paavilainen, and Reinikainen 1989; Jaramillo, Alku, and Paavilainen 1999; Jaramillo et al. 2001; Amenedo and Escera 2000; Menning et al. 2002; Nenonen et al. 2003; Ylinen, Huotilainen, and Näätänen 2005; Ylinen et al. 2005, 2006; Kirmse et al. 2008; Chládková, Escudero, and Lipski 2013) and tone (e.g., Gandour et al. 2000, 2004; Gandour 2006, 2007; Chandrasekaran, Krishnan, and Gandour 2007, 2009; Kaan et al. 2008; Fournier et al. 2010), but there is hardly any research on the interaction of these cues for word processing. Only a few studies explored the processing of fundamental frequency (F0) and length, but only on the basis of isolated sounds (e.g., Czigler and Winkler 1996; Levänen et al. 1993; Wolff and Schröger 2001; Jaramillo et al. 2001) or vowels (e.g., Jaramillo et al. 2001) and therefore without direct relevance to word access. Moreover, a consensus on the existence of an additive processing of different cues (cf. Wolff and Schröger 2001) or a separated processing (cf. Jaramillo et al. 2001) remains to be found.

A number of results from the studies concerning cognitive processing of length or acoustic duration differences are summarised in the following: i) Depending on the direction in which the various stimuli are tested (long to short or short to long), MMN effects diverge to a different degree (cf. Jaramillo, Alku, and Paavilainen 1999; Takegata et al. 2008; Colin et al. 2009). Similarly, ii) the stimulus material used (e.g., harmonic tones vs. vowels) is considered to be another factor which influences the strength of the effect (cf. Jaramillo, Alku, and Paavilainen 1999; Jaramillo et al. 2001; Takegata et al. 2008; Christmann et al. 2014), and also iii) the phonological relevance of length in native languages seems to be a relevant factor (cf. Nenonen et al. 2003; Ylinen et al. 2006; Kirmse et al. 2008). It can be concluded that differences in length represent an important value for neural processing and depend on varying factors. But the stimuli used only range from harmonic tones to individual vowels to pseudowords. So the question remains as to whether the identified neural signatures can also be transferred to real words because differences in phonological quantity have an important function for lexical access unlike pure duration differences between harmonic tones.

The results mentioned above are valid only for simple length contrasts, but in some phonological systems they appear in conjunction with other prosodic features. Therefore, the question arises how quantity differences are processed when additional cues like tone accents can be used for lexical access.

Varieties in which such an issue can be investigated very well are the Moselle-Franconian dialects. These dialects spoken in the western part of Germany are of particular interest to researchers as they have developed a form-meaning mapping in which lexical and morphosyntactic distinctions between words are often derived from prosodic, i.e., tonal cues alone (as referred to tone accents like Tone Accent 1 and Tone Accent 2; cf. Schmidt 1986; Gussenhoven and Peters 2004; Werth 2011; Köhnlein 2011). In the Moselle-Franconian dialects "the contrast is acoustically manifested by a complex phenomenon consisting of a constant length opposition (Tone Accent 1 is always shorter) and a robust pitch difference" (Werth 2012: 187–188). Werth (2012: 192) describes the pitch difference as follows: "Accent 2 is marked with a lexical high tone [H_{lex}] on the second mora which Accent 1 lacks". For example, in the dialect of Mayen, the dialect words [d̥a̠͡ʊ^{1}f] 'pigeon' and [d̥a̠͡ʊ^{2}f] 'baptism' are distinguished by prosodic tone accent features (characterised by superscripts in the transcription) but not by features at a segmental level, while in the Standard German variety the phonemic segmental distinction /b/ versus /f/ is used to differentiate between the same lexemes: [ta̠͡ʊbə] 'pigeon' and [ta̠͡ʊfə] 'baptism'. Roughly speaking, compared to Standard German, Moselle-Franconian dialects show a preferred tendency to distinguish word meaning by prosody in that tone accents occur in all words that include heavy nuclei, i.e., on syllables containing a long vowel, a diphthong or a short vowel followed by a sonorant coda (cf. Schmidt 1986; Werth 2011; Köhnlein 2011).

Taken together, the previous considerations may mean that the combination of certain kinds of prosodic features like tone accents and vowel quantity are more relevant for lexical access in a tone-accent dialect than in the Standard German variety. To address this issue, we have investigated the interaction of two types of prosodic features in natural speech: i) tone-text-association, which is the association of tones with prosodic domains like syllables, moras or words, and ii) vowel quantity. To do so, we tested two participant groups: i) speakers of the Moselle-Franconian dialect who use tone accents as a cue for lexical access and ii) speakers of the Standard German variety who do not.

1.2 The interaction of tone and length in the Moselle-Franconian dialect of Mayen

Length is a feature in the Moselle-Franconian dialects which can be complementarily distributed between vowels and the following sonorant. Its auditory and acoustic salience is clearly evident in words with Tone Accent 2 (cf. Schmidt 1986: 185–191). Thus, if there is a long vowel followed by a sonorant in the rhyme of words with Tone Accent 2, the sonorant is short (e.g., [ʃa̠ː²l] 'stale', see Table 1, condition b.). On the other hand, if the vowel is short, the sonorant is prolonged (e.g., [ʃa̠l·²] 'acoustic noise', see Table 1, condition a. In CV phonology (cf. e.g., Clements and Keyser 1983; Lass 1984; Hayes 1999), these time units can be represented with the skeletal positions VCC (= short vowel + long sonorant) and VVC (= long vowel + short sonorant).[3] Phonetically, the occurrence of long sonorants can be explained by the fact that the realization of Tone Accent 2 requires more time than available in two short consecutive segments (short vowel + short sonorant).[4] By prolonging the sonorant, a sufficiently large time interval for full realization of the tonal pattern is provided (cf. Schmidt 1986: 185–191).

In some phonological theories, this large time interval can be analyzed through an underlying bimoraic domain (μμ), if we follow the definition of moras as tone-bearing units. Often, moras are defined units of weight and length (cf. e.g., Trubetzkoy 1939; Féry 2001; Hyman 2003), but in most tone and tonal accent languages, moras also operate as tone-bearing units (cf. e.g., Zec 1994; Hyman 2003; Zhang 2002; Werth 2011, 2012), which is their crucial function within the Moselle-Franconian dialects. Both moras are of decisive importance; intonational tone associated with the first mora expresses the communicative meaning, while the pitch level on the second mora (presence of a lexical high tone for Tone Accent 2 and absence for Tone Accent 1) expresses the lexical meaning (cf. Werth 2011, 2012). Although sonorant consonants are lengthened due to the association of the tone accent with the second mora, consonantal length alone could not be seen here as a phonologically distinctive feature. Schmidt (1986) assumes that

3 Other authors like Wiese (1996), Ramers (1998) or Ramers and Vater (1995) associate also long vowels with two skeletal positions, but with the symbols VC. To reflect the difference between the two mentioned conditions (long vowel + short sonorant consonant and short vowel + long sonorant consonant), we follow authors like Lass (1984) and Hayes (1999), using the symbols VV for long vowels in the phonological representation.

4 This larger time frame required for the realization of high tones in opposition to low tones is due to the fact that an increase in fundamental frequency in speech production universally requires more muscle activity and therefore more production time than for an F0-decrease (cf. Ohala 1972, 1978; Ohala and Ewan 1973; Sundberg 1973, 1979).

phonetic length of the sonorant consonant is to be regarded solely as a suprasegmental feature expression of Tone Accent 2. Thus, when preceded by a short vowel, the sonorant is lengthened to provide enough space for the realization of tonal information on the second mora. Therefore, length differences on the sonorant consonants are an epiphenomenon of the association of Tone Accent 2-lexical high tone with the consonant.

Another study that deals with Moselle-Franconian tone accents, and thus also constitutes direct relevance for the present study, is by Werth (2011, cf. 2012). Acoustically, tone accents are based on a phonetic-prosodic feature complex of fundamental frequency, duration and intensity, as has been demonstrated in numerous studies (cf. Heike 1962; Schmidt 1986; Gussenhoven and Peters 2004; Peters 2006; Werth 2011). However, Werth (2011) has shown that for the identification of tone accents, the tonal features represent the crucial cue in perception. In his perception tests with native speakers, he manipulated all of the three acoustic components belonging to tone accents and came to the result that “native listeners identified the opposite tone accent (Tone Accent 2 in place of Tone Accent 1 and vice versa) significantly often when F0 was manipulated, but almost never when length [...] or intensity [...] were” (Werth 2012: 190). Thus, length can be considered a redundant feature, whereas the lexical tone represents the relevant information for the identification process.

Another important result of the identification task was that listeners always responded to the F0-movement at the end of the word. For this reason, it seems sensible to associate the lexical tone to the second mora. Furthermore, this lexical tone is always associated with a long segment with an underlying bimoraic structure; in monosyllabic words with a VCC-structure (condition a.: short vowel + long sonorant consonant), the second mora and therefore the lexical high tone (H_{lex}) is located on the consonant. Thus, a tone accent cannot be identified until the sonorant has been perceived. In monosyllabic words with a VVC-structure (long vowel + short sonorant consonant), the second mora is already located on the vowel and can be identified as soon as the vowel has been perceived.

Thus, tone accent speakers use a combination of several cues (vowel quantity, tone-text-association and length of the sonorant consonant) for lexical access; however, according to the results of Werth (2011), length appears to be redundant while tone is relevant. However, there are minimal pairs with contrasts in vowel quantity in the Moselle-Franconian dialects, as well (e.g., [ʃa̱lˑ^{2}] ‘acoustic noise’ – [ʃa̱ː^{2}l] ‘stale’). Thus, the question remains whether tonal cues are redundant and less relevant in online word recognition than vowel quantity in these word pairs.

In sum, the phonetic and phonological conditions of monosyllabic words contrasting in vowel quantity, tone-text-association and lateral length and the relevance of these cues for the perception of Moselle-Franconian dialect speakers can be represented as follows:

Tab. 1: Phonological and perceptual representation of monosyllabic words with inverted length contrasts combined with Tone Accent 2 in the Moselle-Franconian dialect of Mayen

Condition	**a.**			**b.**		
Phonology	μ		μ	μ		μ
	V		CC	VV		C
	short	+	long^{Hlex}	long^{Hlex}	+	short
Perception	ø		2 cues (H_{lex}+ length)	H_{lex}	+	length

1.3 The interaction of tone and length in the Standard German variety

As tone accents do not occur in the Standard German variety, the interaction of tone and length information must be different from that of the Moselle-Franconian. Tonal information is not used to differentiate lexical meanings, but is processed in another way. Monosyllabic stimuli realised with Tone Accent 2 by a Moselle-Franconian speaker (see Table 2, condition c. and d.) can be processed by the Standard German listeners, as well. But without dialectal competence they are only able to perceive a simple high tone (H*) without further lexical meaning. In addition to the variation of loudness in Standard German, changes in the temporal structure as well as pitch changes are used to realise stress (cf. Stock and Zacharias 1982; Jessen et al. 1995; Dogil 1999; Pompino-Marschall 2003 etc.). Therefore, there may be a certain sensitivity towards the perception of these prosodic parameters. This means that the length of vowels and consonants as well as the high tone are perceived, but only vowel quantity is used distinctively for lexical retrieval with the beginning of the lateral.

Vowel length is a phonological cue in the Standard German variety with low functional load. Unlike the Moselle-Franconian dialects or real quantity languages such as the Finno-Ugric languages Finnish, Hungarian or Estonian, most quantity contrasts in Standard German are coupled with differences in vowel quality, e.g., /oː/ – /ɔ/ or /iː/ – /ɪ/. The only exceptions are the vowel pairs /ɛ/ –

/ɛː/ and /a̠/ – /a̠ː/, which differ only in the feature [± long] (cf. Fuhrhop and Peters 2013). Long consonants with sense-discriminative function like the Finno-Ugric languages or long sonorant consonants due to lexical tones like the Moselle-Franconian dialects do not exist in the Standard German variety.

In sum, it can be stated that there is a difference in the phonological prosodic systems between Moselle-Franconian dialect speakers and speakers of the Standard German variety concerning the relevance of tone accents and of contrasts in length for lexical access. Vowel quantity has a low functional load in the Standard German variety as well, but there is no phonological relevance of tonal information and consonantal length.

The present study will examine the question how important the various cues are in the neural processing of the two mentioned participant groups. According to the results of the perception study by Werth (2011), it can be deduced that there must be a hierarchy for the individual cues in processing, i.e., primarily and secondarily relevant cues. Commonly occurring cues like the "duration ratio" (vowel + closure), formant transition and the length of voicing are used (with varying relevance) for the identification of fortis and lenis in the context of nasal plosion in the Standard German variety (cf. Kohler 1979). Gussenhoven (2004) describes this phenomenon as contrast enhancement that is defined for phonology in Werth (2012: 196) as "the fact that linguistic function (in a broader sense) is encoded in several different formal dimensions". In our study, we expect visible differences in the hierarchy of the various cues because of their differing phonological relevance.

Thus, the phonetic and phonological conditions of monosyllabic words contrasting in vowel quantity, tone-text-association and lateral length as well as the relevance of these cues for the perception of Standard German speakers can be represented as follows:

Tab. 2: Phonological and perceptual representation of monosyllabic words with inverted length contrasts combined with a high tone in non-dialectal listening competence

Condition	**c.**			**d.**		
Phonology	μ		μ	μ		μ
	V		CC	VV		C
	short	+	long^{H*}	long^{H*}	+	short
Perception	ø		2 cues (H*+ length)	H*	+	length

2 Methods

To test our hypotheses, we conducted an event-related potential (ERP) study with a categorical oddball paradigm using the electroencephalograpy (EEG) technique. This design was chosen in order to examine the so-called mismatch negativity (MMN), a fronto-central negative component, usually peaking at 150 to 250 ms from change onset. This component is elicited when an infrequent deviation (‘deviant’) occurs among frequently repeated sound patterns (‘standard’). The repetitive presentation of standards creates a short-term memory trace in the auditory cortex for this pattern. If this pattern is violated by an infrequent deviant, an MMN is elicited, as it reflects an automatic, pre-attentive response to any change in auditory stimulation, regardless of the participants’ attention. It thus indicates that this stimulus deviates from the memory representation of the preceding series of standards (cf. Näätänen et al. 2007).

2.1 Stimuli

Two monosyllabic words from the Moselle-Franconian tone accent dialect spoken in Mayen were selected for this study: [ʃa̠ː^{2}l] ‘stale’ – [ʃa̠lˑ^2] ‘acoustic noise’.[5] These words differ in vowel quantity in one phoneme only ([a̠ː] vs. [a̠]), but they have an additional difference in lateral length ([l] vs. [lˑ], see Figure 1), which is caused by the tone-mora-association. Furthermore, there is a difference in the tone-text-association which means that the lexical high tone of Tone Accent 2 is associated with the lateral in [ʃa̠lˑ^2] while in [ʃa̠ː^{2}l] the lexical high tone is already associated with the vowel (cf. Schmidt 1986; Werth 2011 and Figure 1). Together with the minimal pair [ʃa̠l^1] ‘(apple) skin’ and [ʃa̠ˑ^{1}l] ‘shell, shawl’, a total of four lexemes can be distinguished by length differences in combination with tone accent differences (cf. Schmidt 1986). Notice that in the present study, these words with Tone Accent 1 were not tested, only the two lexemes with Tone Accent 2.

5 The tested word pair was part of an experiment in which two other minimal pairs with other phonological contrasts (Tone Accent 1 vs. Tone Accent 2 and vowel quality /o/ vs. /u/) were also examined (cf. Schmidt 2017).

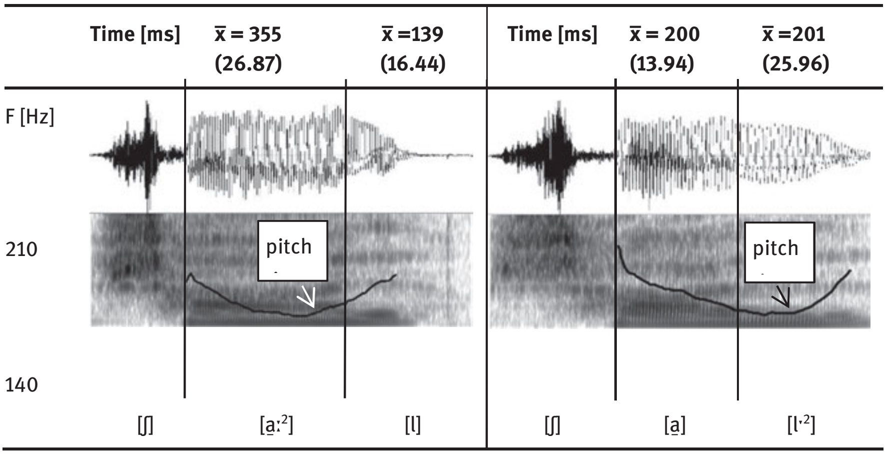


Fig. 1: Average vowel and lateral segment durations and pitch curves of Accent 2 in [ʃa̠ː^{2}l] and [ʃa̠lˑ^{2}]

The critical word pairs were recorded several times to obtain different tokens with a natural variation for each word. Finally, eight different natural tokens per word served as frequently presented stimuli (= standard[6]) and as infrequently presented stimuli that are incongruent with the memory representation of the preceding stimuli, cf. Näätänen et al. (2007) (= deviant). This acoustic variability was chosen to create a more natural speech perception condition as well as a memory trace for the presented standard condition, since it has been shown that a higher and hence more natural variability in standard items leads to a more reliable abstraction or trace form of the different acoustic stimuli presented (Phillips et al. 2000). Moreover, the standard and the deviant representations of one critical item were phonetically identical, i.e., the standards as well as the deviants presented comprised the same eight different tokens. As a result, purely acoustic effects between standards and deviants could be distinguished from phonetic and phonological effects (cf. Scharinger, Monahan, and Idsardi 2011).

All stimuli were digitally recorded with a sampling rate of 44.1 kHz and a 16 bit (mono) sample size, using an electret microphone (Beyerdynamic MC 930) and the software Adobe Audition 2.0. The stimuli were spoken naturally by a female native speaker of the Moselle-Franconian dialect from Mayen at a normal speech rate.

6 If we refer to the Standard German variety in the present study, we always use the term Standard German.

For each word, eight acoustically variant tokens were selected. Table 3 depicts mean values of the parameters pitch, and frequencies of the first two vowel formants for tokens of the two word types. Statistical analyses utilizing Mann-Whitney-U tests revealed that the distribution of pitch values for the two word types differed significantly ($U = 64$, $p < .001$). Despite these statistical differences, it is questioned here that the acoustic differences exceed the perceptional threshold. According to Nooteboom (1997: 645) pitch differences between stimuli can be reliably discriminated, if the pitch difference exceeds a difference of three semitones (one semitone corresponds to a frequency difference of approximately 6%). Consequently, a perceptible difference should exhibit a difference of roughly 25 Hz. The pitch difference observed in our stimulus material is clearly smaller.

Tab. 3: Acoustic properties of conditions with mean values (and standard deviation)

Stimulus	Pitch [Hz]	1. Vowel Formant	2. Vowel Formant
[ʃa̠ː²l]	158 (5.66)	710 (79.55)	1352 (79.72)
[ʃa̠lˑ²]	169 (8.57)	753 (28.59)	1365 (80.46)

Finally, all items were controlled for and normalised in intensity to ~ 80 dB SPL (Table 3). This adjustment was carried out using the sound recording and analysis software PRAAT (version 5.3.08, Boersma and Weenink 2012).

2.2 Procedure

The experiment was carried out twice: first with participants from Mayen raised in the Moselle-Franconian dialect area and in a second run with Standard German speakers in Marburg. In both group sessions, exactly the same conditions and the identical set of stimuli were used. All stimuli of the minimal pair [ʃa̠ː²l] – [ʃa̠lˑ²] were presented in two experimental blocks. In one block, stimuli with a long vowel served as standard and stimuli with a short vowel as deviant. The opposite was the case in another block in which a short vowel served as standard and a long vowel as deviant. In total, together with two further conditions not reported here,[7] six blocks consisting of 1000 items each (15% deviants) were presented with each requiring approximately 25 minutes.

In order to avoid sequence effects, the block order was varied across participants. Moreover, two blocks containing the same lexical material were never presented directly one after the other. Each block started with ten standards which were not included in data analysis. Next, the standards and deviants were presented in a classic passive oddball paradigm, i.e., in a pseudo-randomised order in which a deviant was presented after two up to eight standards. The inter-stimulus interval (offset-to-onset) was 900 ms. Stimuli were presented via two loudspeakers at a comfortable listening level (~ 65 dB SPL).

During the experiment, participants were comfortably seated in front of a computer screen in a dimly lit and quiet room. They were instructed to watch a silent movie and to disregard the auditory presentation. After the first block, all participants reported that they were able to ignore the auditory signal and to concentrate entirely on the movie presented. Between blocks, participants were offered a break to rest their eyes. All procedures were performed in compliance with relevant laws and institutional guidelines.

2.3 Participants

25 native speakers of the Moselle-Franconian dialect from Mayen (16 women; mean age 50.0,[8] age range 31 to 62) and 21 speakers of Standard German not born or raised in the Moselle-Franconian dialect area (16 women, mean age 52.8, age range 42 to 61) participated in the two distinct experimental sessions. Both groups were closely matched in their mean age. All subjects were right-handed, monolingual and had normal or corrected-to-normal vision. All participants with a dialectal background were born and raised in Mayen and still live there. Their dialect competence was tested and verified via a dialect pre-test. None of the participants had hearing deficits, which was verified by an online-hearing-test.[9]

7 Two other minimal pairs ([ʃtuf] 'living room' – [ʃtof] 'textile fabric' and [d̥a̠͡ʊ¹f] 'pigeon' – [d̥a̠͡ʊ²f] 'baptism') were tested in the same experiment beside the pair with contrast in vowel quantity presented in this study (cf. Schmidt 2017).

8 The mean age range for ERP studies is typically between 18 and 35 years. The relatively high mean age in both participant groups in the present study is due to the factor of dialect competence in the dialect group. Dialect competence is generally more stable in older peer groups so that the age range had to be extended for the purposes of the present study.

9 URL: http://www.powerone-batteries.com/de/wissen/hoertest/power-one-hoertest/?no_cache=1 (accessed 20 May 2016).

All participants gave their informed consent to this study and privacy rights were thoroughly obeyed. Each participant received monetary compensation for taking part in the study.

2.4 ERP recording and data processing

An electroencephalogram (EEG) was recorded from 26 Ag/AgCl electrodes, mounted on an elastic cap (EasyCap), according to the 10-20 system (F7, F3, Fz, F4, F8, FC5, FC1, FCz, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CPz, CP2, CP6, P7, P3, Pz, P4, P8, POz) with a "BrainVision" (Brain Products GmbH) amplifier. The C2 electrode served as the ground electrode. The reference electrode was placed at the tip of the nose. Four electrodes measured the **electrooculogram (EOG)**, i.e., horizontal and vertical eye movements to control for eye movements and blinks. Two electrodes were placed at the left and right mastoid. EEG and EOG were recorded continuously with a sampling rate of 500 Hz and filtered offline with a 0.16 to 30 Hz bandpass filter. All electrode impedances were kept below 5 kΩ. EEG recordings were re-referenced off-line to the linked mastoids to decrease the signal-to-noise ratio and hence to increase the MMN amplitude (cf. Schröger 1998). Averaged data were baseline corrected over 100 ms before vowel onset.

For the data analysis, all standard and deviant epochs starting at a baseline of 100 ms before the divergence point up to 900 ms after the vowel onset were automatically scanned for artifacts produced by eye or body movements. All epochs that included artifacts with an amplitude exceeding 75 microvolt were removed from the data set. Subsequently, all single-trial waveforms were individually screened for further artifacts. As a result of these observations, the data sets gathered from ten participants (eight women) in the dialect group and six participants (six women) in the Standard German group had to be excluded from the analysis because of the high number of artifacts (mainly eye blinks) per condition epoch (more than 50%). Thus, the data sets of 15 participants[10] were analysed per group (Moselle-Franconian dialect group: eight women, mean age 50.7 years; Standard German group: ten women, mean age 53.1 years). ERP responses to the first ten standard stimuli of each experimental block as well as to standard epochs immediately preceded by a deviant were excluded from data analysis.

10 Due to the given high mean age and its known influence on ERP components it would have been interesting to test two age groups per participant group to control for this factor. However, due to the small groups, a division into two age groups within each group of participants is not possible. We therefore refer to research on age effects on ERP patterns in the following discussion.

2.5 Data analyses

For the statistical analysis, an omnibus multifactorial repeated measures ANOVA was calculated with the factor GROUP (Moselle-Franconian dialect participants vs. Standard German participants), REGION (frontal [F3, Fz, F4], central [C3, Cz, C4], parietal [P3, Pz, P4]) and VOWEL QUANTITY (long vowel vs. short vowel). Averages were calculated from the onset of the divergence point up to 900 ms thereafter, with a baseline of 100 ms. For the statistical analysis, consecutive epochs of 50 ms were investigated in both groups in the classical MMN time window (100 to 200 ms).

Furthermore, additional time windows were calculated following visual inspection of the grand-average curves within each group. For these additional time windows, an omnibus ANOVA with the factors described above as well as a multifactorial ANOVA within each group was conducted. For the latter analyses, time windows were adjusted on the basis of visual inspection of the grand average curves. This was necessary as the effects' latency differed between the two groups. This was expected because of the varying dialect background between the two groups. For effects with more than one degree of freedom, Huynh-Feldt (1976) corrections were applied and corrected p-values are reported here.

3 Results

Data analyses aim at finding mean voltage differences between certain experimental conditions that manifest differently in the two experimental groups. In particular, the oddball paradigm leads to the expectation to find a significant negative mismatch response for deviants indexing pre-attentive perception of duration contrasts. Table 2 shows that negativity effects were found in both participant groups. Note that all presented results are derived from comparisons across oddball blocks, i.e., deviant related effects were examined by comparing phonetically identical stimuli. For example, [ʃa̠ː²l], which was presented as deviant in one block, was compared with [ʃa̠ː²l], which was presented as standard in another block. This procedure ensured that potential effects between standard and deviant which are purely elicited by acoustic differences could be excluded (cf. Eulitz and Lahiri 2004; Scharinger, Monahan, and Idsardi 2011).

For the comparison of amplitudes of the standard condition [ʃa̠ː²l] and the deviant condition [ʃa̠ː²l] in the time window between 100 and 200 ms, the omnibus ANOVA showed significant main effects for all three factors GROUP [$F(1, 28) = 4.81$, $p = .037$, $\eta^2p = .11$], REGION [$F(2, 56) = 6.53$, $p = .010$, $\eta^2p = .02$] and VOWEL QUANTITY

[$F(1, 28) = 4.94$, $p = .034$, $\eta^2p = .02$]. There was no significant three way interaction, but a significant interaction between the factors GROUP and REGION [$F(2, 56) = 8.80$, $p = .003$, $\eta^2p = .03$]. This significant interaction is in line with the expectation to find a more pronounced negativity in the frontal region as the MMN is typically distributed frontally. In fact, the post-hoc analysis of this interaction by REGION only revealed a significant result in the frontal region [frontal: factor GROUP: $F(1, 28) = 8.28$, $p < .05$, $\eta^2p = .20$; factor VOWEL QUANTITY: $F(1, 28) = 5.71$, $p < .05$, $\eta^2p = .03$].

In order to also resolve this significant interaction by GROUP and because of the significant main effects for all three factors, a multifactorial repeated measure ANOVA was also conducted for this time window within each group. This should have revealed possible between-group differences in the early MMN time window. However, these within-group calculations showed no significant main effects for the main factor VOWEL QUANTITY.

The comparison of the short vowel contrast between the standard condition [ʃa̠lˑ^{2}] and the deviant condition [ʃa̠lˑ^{2}] in the time window from 100 to 200 ms only elicited a main effect for REGION [$F(2, 56) = 29.12$, $p = .000$, $\eta^2p = .01$]. However, in contrast to the typical early MMN time window, significant mismatch effects were found in later time windows for both contrast pairs. The comparison of the long vowel contrast between the standard condition [ʃa̠ː^{2}l] and the deviant condition [ʃa̠ː^{2}l] in a time window from 300 to 450 ms revealed significant main effects for the factors GROUP [$F(1, 28) = 9.13$, $p = .005$, $\eta^2p = .19$], REGION [$F(2, 56) = 14.91$, $p = .000$, $\eta^2p = .07$] and VOWEL QUANTITY [$F(1, 28) = 11.32$, $p = .002$, $\eta^2p = .05$].

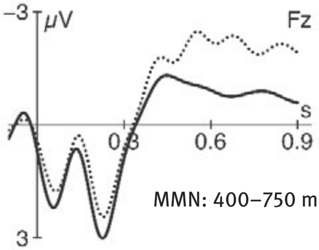
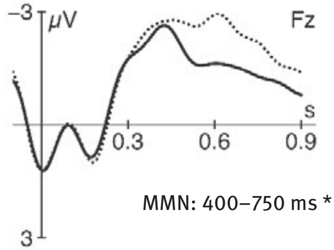
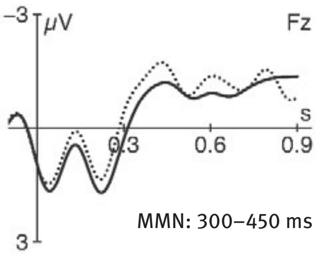
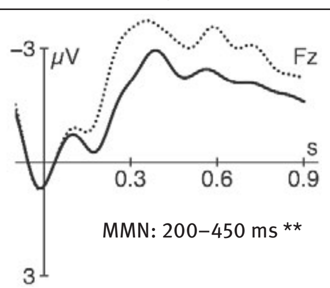
There was no significant three way interaction, but a significant interaction between the factors GROUP and REGION [$F(2, 56) = 8.06$, $p = .006$, $\eta^2p = .04$]. The post-hoc analysis of this interaction by REGION revealed a more pronounced effect in the fronto-central region [frontal: factor GROUP: $F(1, 28) = 10.11$, $p < .01$, $\eta^2p = .24$; factor VOWEL QUANTITY: $F(1, 28) = 9.53$, $p < .01$, $\eta^2p = .05$; central: factor GROUP: $F(1, 28) = 8.78$, $p < .01$, $\eta^2p = .21$; factor VOWEL QUANTITY: $F(1, 28) = 10.29$, $p < .01$, $\eta^2p = .05$]. A post-hoc within-group comparison revealed a significant latency difference for the negativity effect between the two groups: it starts 100 ms earlier in the Standard German group than in the dialect group.

Due to this latency difference, post-hoc analyses were calculated for each group with the factors VOWEL QUANTITY and REGION in a time window from 300 to 450 ms for the dialect group and from 200 to 450 ms for the Standard German group. These analyses show significant main effects in both groups for the factor VOWEL QUANTITY [dialect group: $F(1, 14) = 6.04$, $p = .028$, $\eta^2p = .04$ / Standard German group: $F(1, 14) = 9.50$, $p = .008$, $\eta^2p = .09$] but no significant interaction between the factors VOWEL QUANTITY and REGION. The measurements of the deviant peak amplitude and peak latency also show that the negativity effect is more pronounced and peaks earlier in the Standard German group in comparison to the peak latency and peak amplitude of the deviant in the dialect group (see Table 5 in the appendix for peak latencies and amplitudes of the late MMN at Fz for all experimental conditions).

For the comparison of the short vowel contrast between the standard condition [ʃa̠lˑ^{2}] and the deviant condition [ʃa̠lˑ^{2}], similar results were found. For the omnibus ANOVA, a time window between 400 and 750 ms was investigated. It revealed significant main effects for the factors REGION [$F(2, 56) = 20.35$, $p = .000$, $\eta^2p = .08$] and VOWEL QUANTITY [$F(1, 28) = 22.58$, $p = .000$, $\eta^2p = .10$] but no significant interactions. However, hypothesis guided post-hoc analyses within each group were calculated from 400 to 750 ms for both groups. For the dialect group, a significant main effect was found for the factor VOWEL QUANTITY [$F(1, 14) = 14.24$, $p = .002$, $\eta^2p = .13$] without a significant interaction between the two factors VOWEL QUANTITY and REGION. Within the Standard German group, significant main effects were elicited for the factors REGION [$F(2, 28) = 21.14$, $p = .000$, $\eta^2p = .13$] and VOWEL QUANTITY [$F(1, 14) = 8.72$, $p = .010$, $\eta^2p = .08$], but again no significant interaction between these factors was found. For this contrast in vowel quantity, no significant latency differences were found between the two participant groups.

Table 4 displays an overview of the relevant time windows and the statistically significant results in both groups of participants for the critical word pair [ʃa̠ːˑ^{2}l] – [ʃa̠lˑ^{2}].

Tab. 4: Grand average ERP responses for deviants (dotted lines) and standards (solid lines) at Fz of the two experimental conditions [ʃa̠l·²] (a.+c.) and [ʃa̠ː²l] (b.+d.) in different time windows for both groups of participants, measured from 100 ms prior the vowel onset up to 900 ms. Statistical significance is indicated by * ($p < .05$), ** ($p < .01$), *** ($p < .001$).

	Dialect speakers	**Standard German speakers**
Condition	a.	b.
Phonology	μ μ V CC short + long^Hlex	μ μ V CC short + long^H*
Perception	ø 2 cues (Hlex+ length)	ø 2 cues (H*+ length)
EEG-Responses Legend: ······· deviant —— standard	 MMN: 400–750 ms **	 MMN: 400–750 ms *
Condition	c.	d.
Phonology	μ μ VV C long^Hlex + short	μ μ VV C long^H* + short
Perception	Hlex + length	H* + length
EEG-Responses Legend: ······· deviant —— standard	 MMN: 300–450 ms *	 MMN: 200–450 ms **

4 Discussion

4.1 ERP component discussion

While differences in tone-text-association have not yet been in the focus of neurolinguistic research interest, length differences in pre-attentive processing regarding isolated vowels, pseudowords and tones have already been well examined (cf. Kirmse et al. 2008; Ylinen et al. 2006; Nenonen et al. 2003; Jaramillo, Alku, and Paavilainen 1999; Jaramillo et al. 2001).

In the present study, we used real words contrasting in vowel quantity and found comparable negativities, albeit with differences in terms of their respective latency. The component that was observed in the former studies was classified as an MMN (= Mismatch Negativity). It usually occurs if a mismatch between a memory trace of the currently valid auditory representation and an appearing stimulus takes place (cf. Schröger, SanMiguel, and Bendixen 2013). The latency for this component is usually reported between 100 to 250 ms after the beginning of an alteration (cf. Schröger, SanMiguel, and Bendixen 2013). If we look at our results, the latency of the detected negative-going deflection occurs at a later time window in most cases. Nevertheless, it seems unlikely that this component could be interpreted differently from previous similar studies. Although the latency is similar to the N400 component, the negativity effect found is not distributed largest over centro-parietal sites, which is usually described for the N400 (cf. Lau, Phillips, and Poeppel 2008; Kutas and Federmeier 2011), but has the fronto-central topographic distribution of the MMN.

Furthermore, the experimental design containing an oddball paradigm without an active task leads us to interpret the observed component as an expression of pre-attentive processing. From other studies, it is already known that the MMN may sometimes appear in a later time window, if complex auditory or even linguistic stimuli were used (cf. Korpilahti, Lang, and Aaltonen 1995; Korpilahti et al. 2001; Cheour et al. 2001), as is the case in the present study. Korpilahti et al. (2001) associate this late MMN (lMMN) with the automatic detection of lexical differences. In our study, we investigated phonological quantity differences using the words [ʃa̠ː²l] ‘stale’ and [ʃa̠lˑ²] ‘acoustic noise’, which also result in lexical differences. Thus, this could be an explanation for the difference in latency compared with length contrasts in isolated vowels, pseudowords or tones.[11]

11 Furthermore, it must be taken into consideration that the two groups of participants in our study are older than the usually younger subjects in electrophysiological experiments (mean age in our study: dialect group = 50.0, Standard German group = 52.8). Age effects are also found in

As an initial result of our study, we conclude that phonological quantity contrasts embedded in real words lead to a similar negative component in pre-attentive processing for dialect as well as for non-dialect speakers, similar to length contrasts without lexical access, but with a delay in latency due to the higher complexity of the stimuli.

4.2 Participant group 1 (Moselle-Franconian dialect speakers)

As is known from other studies, MMN effects are highly dependent on the direction of stimulus presentation in their amplitude and latency (cf. Jaramillo, Alku, and Paavilainen 1999; Takegata et al. 2008; Kirmse et al. 2008; Colin et al. 2009). In the present study we find differences as well: a decreased latency of the MMN component (a. vs. b.) and an earlier peak in the deviant condition [ʃa̠ː²l] (see Table 4, cf. Table 5 in Appendix). A closer look at the structure of the stimuli and the position of the high tone may deliver an explanation. While the words in both conditions [ʃa̠ː²l] and [ʃa̠lˑ²] bear Accent 2, the tone-text-association is different. Accent 2 bears the high tone on the second mora, thus the tone accent is associated with the vowel in [ʃa̠ː²l] as described in Section 2.1. In [ʃa̠lˑ²], the vowel bears no tonal information because the second mora is associated with the lateral. Consequently, the final lateral bears the tonal information. Thus, listeners are already able to perceive the difference through the tone-text-association when they hear the long vowel in the condition [ʃa̠ː²l]. Since the vowel quantity can be identified only after the offset of the vowel (cf. Czigler and Winkler 1996; Levänen et al. 1993), the processing of this cue can only start at the beginning of the lateral.

In the condition [ʃa̠lˑ²], the determination of **both** cues can only be accomplished by the perception of the lateral. Because the vowel has to be completed for vowel quantification and the tonal information is tied to the lateral, the listener has to wait for the lateral to gain lexical access. Thus, the vowel itself carries no information for the condition [ʃa̠lˑ²] that could be perceived as different from the standard stimuli. In the condition [ʃa̠ː²l], a negative component with an earlier latency seems to be triggered through the association of the high tone with the vowel. The results found suggest that lexical retrieval already starts with the perception of the high tone when this information is available before the quantity

pre-attentive processing with regard to an ERP's latency and distribution. This is especially reflected in smaller amplitudes (cf. Verleger et al. 1991; Pekkonen et al. 1996; Bertoli, Smurzynski, and Probst 2002; Cooper et al. 2006; Schiff et al. 2008; Rimmele et al. 2012), but also in a later appearance of the MMN in older participant groups (cf. Verleger et al. 1991; Bertoli, Smurzynski, and Probst 2002; Cooper et al. 2006).

cue. This suggests that vowel quantity is only a minor cue for lexical access in the Moselle-Franconian dialect.

Furthermore, the higher amplitude in the condition a. [ʃa̠lˑ²] (see Table 4, cf. Table 5 in Appendix) indicates that the short vowel evokes an increased neural activity in pre-attentive processing. This result is in line with intrinsic stimulus effects presented in studies by Jaramillo, Alku, and Paavilainen (1999), Colin et al. (2009) and Sittiprapaporn (2012). Based on synthetically generated sound patterns, Colin et al. (2009) found a major effect in MMN responses in deviant sound patterns with shorter durations which evoked higher amplitudes than those of longer duration. These results suggest a general effect of vowel length on speech processing. The length effect on MMN amplitudes can be explained by the delay between the moment of deviance detection and the end of the deviant quantification process. In short deviants, deviance detection and quantification take place simultaneously, giving rise to the MMN amplitude. By contrast, in long deviants, deviance detection and quantification occur separately one after another. However, the effects observed here cannot be explained solely by intrinsic effects since the amplitude effects for the quantity contrast are highest for the tested quantity contrast in comparison to other contrasts (vowel quality and tone contrasts) tested in the same experiment. While the tested contrasts in tone and vowel quality were single deviants, the examined word pair in the present study is characterised through three differing features (tone-text-association, vowel quantity and lateral length). This difference might suggest that all these features are processed additively in the condition [ʃa̠lˑ²], in which all occur during a short period of time, more precisely during the lateral perception. While vowel quantity could only be identified with the beginning of the lateral and lateral length with the end of this sonorant consonant, even the third cue, the high tone, is associated with the lateral. Thus, all cues are perceived during the 201 ms (mean value) of the lateral, which leads to an additive processing.

According to Ylinen, Huotilainen, and Näätänen (2005), additive responses indicate an independent analysis of the different features. The authors examined in their pseudoword MMN study the processing of stimuli that differ in terms of quality ([ipːi]), quantity ([iti]) or a combination of both ([ipi]) from a standard ([itːi]). By comparing the MMN results for single and combined deviant forms, they found that both features are processed additively and therefore the analysis of these features is carried out independently.

In the present analysis, the nature of the stimuli is slightly different; nevertheless, in the condition [ʃa̠ː²l], deviating features occur one after another in an average time slot of 316.5 ms (beginning of the high tone at the middle of the vowel $\bar{x}$ = 177.5 ms (see Figure 1), vowel quantity with the start of the lateral and lateral length with the end of the lateral $\bar{x}$ = 139 ms), while they are perceptible within one segment ([lˑ^{2}]) in the condition [ʃa̠lˑ^{2}]. The higher amplitude in condition [ʃa̠lˑ^{2}] could thus be an indication that not only phoneme quality and quantity, as described by Ylinen, Huotilainen, and Näätänen (2005) is processed independently, but also the two suprasegmental features of quantity and tone accents, if they both occur at the same time or in a very short period of time.

4.3 Participant group 2 (Standard German speakers)

As mentioned before, the main difference between the two participant groups of dialect and Standard German speakers is the phonological relevance of the tone accent features like the tone-text-association. However, there are some similarities in processing the investigated word pair with regard to latency and amplitude differences between the MMNs for both deviants.

Both deviants elicited significant late MMN effects in the Standard German speaker group, as well. Moreover, an earlier latency in the deviant condition [ʃa̠ː^{2}l] contrasting with [ʃa̠lˑ^{2}] (a. + c. vs. b. + d., see Table 4) can be observed in both participant groups. Thus, although high tones are not associated with lexical relevance in Standard German, Standard German speakers perceive this tonal information at different temporal points in both presented words. We therefore conclude that the high tone, regardless of whether it is used for lexical access (dialect group) or only as a salient marker (Standard German group), seems highly relevant in pre-attentive processing of deviance.

Furthermore, the higher amplitude in the condition c. [ʃa̠lˑ^{2}] vs. d. [ʃa̠ː^{2}l] (see Table 4, cf. Table 5 in Appendix) indicates that the short vowel with Accent 2 associated with the lateral evokes an increased neural activity in pre-attentive processing. Even without phonological relevance of the tone-text-association, additive processing of the two suprasegmental features quantity and tone accent takes place nearly at the same time. It seems like Standard German speakers perceive the high tone as a salient cue, and add this information to vowel quantity resulting in increased processing costs in comparison to the condition where these two cues are processed successively (condition d. [ʃa̠ː^{2}l], see Table 4, cf. Table 5 in Appendix). Thus, we conclude that tonal information is processed independently from quantity regardless of phonological function.

Although the negativity effects occur in both groups, there are some relevant differences between the groups regarding the latency of the effects (cf. b. vs. d., see Table 4). For deviants with a long vowel, the effect starts 100 ms later in the dialect group. We assume that one reason for this difference in latency is the tonal information of Accent 2, as it represents an extra processing effort for the participants who speak the Moselle-Franconian dialect. Accent 2, characterised by a final rising pitch or the presence of a high tone on the second mora (see Introduction), plays an important role as a cue for lexical access during vowel perception, but only for speakers of the Moselle-Franconian dialect. While in Standard German no lexical meaning is carried by tone accents, Moselle-Franconian dialect speakers use this information to disambiguate lexemes and, according to our results, to distinguish vowel quantity, if the high tone precedes the vowel offset. With the high tone on the second mora, Moselle-Franconian dialect speakers may predict that the similar lexemes [ʃa̠ˑ¹l] 'shell, shawl' and [ʃa̠l¹] '(apple) skin', which lack this high tone on the second mora due to Accent 1 (cf. Werth 2011), can be excluded from the retrieval process. Since the second mora is associated with the vowel, the lexeme [ʃa̠lˑ²] 'acoustic noise' can also be excluded because here the high tone is located on the lateral. Thus, the tone accent initiates lexical access while the length of the vowel ensures the identification for the group of dialect speakers. Since only the group of dialect speakers are able to use the tonal information for lexical processing and word recognition, their cognitive effort is even higher compared to Standard German speakers who have to process the quantity cues only for lexical retrieval. For Standard German speakers, the recognition of the high tone results only in the perception of differing phonetic information or the marking of salience or word stress without any further distinction. Thus, the high tone is an important lexical cue for the dialect group only. This higher complexity in processing might delay the negativity onset compared to the Standard German group.

Moreover, due to the later onset of the negativity effect, its latency is generally shorter than the effect elicited in Standard German speakers. Shorter MMN latencies have often been associated with higher discrimination sensitivity (cf. Kirmse et al. 2008; Partanen et al. 2011). Since tonal cues are only lexically relevant for the dialect speakers and the particular phonetic form of the stimuli presented is only known by the Moselle Franconian dialect speakers and not by the Standard German speakers, these advantages might explain the higher discrimination sensitivity of dialect speakers. Because they are used to discriminate tone accent contrasts in their own dialect, they acquire a more pronounced discrimination sensitivity and accuracy. In contrast, the group of Standard German speakers might process the given tonal cue as word stress information, leading to prolonged processing.[12] This process seems more time-consuming because a more complex analysis for the whole word is needed (cf. Kirmse et al. 2008).

Our results thus demonstrate that the quantity contrast combined with tonal information elicits a late MMN in both groups of participants. However, significant latency differences of this component between the two groups reflect their difference in processing a tone accent cue which contains phonologically relevant information only for the participants with competence in the Moselle-Franconian dialect, but not for the Standard German speakers.

5 Conclusion

In the study presented here, we investigated the importance of high tone information associated with different sound segments during the processing of real words with a phonological contrast in vowel quantity (Moselle-Franconian Accent 2 in words with long and short vowels). Our results show a late MMN effect for both conditions (long vs. short vowel) and in both participant groups (Standard German and dialect speakers). Latency differences reflect clear differences between the two groups and their differing phonological systems. Since the high tone carries no lexical meaning for the Standard German speakers, it can be processed as a simple salient intonation contour or a stress cue. In contrast, the dialect speakers use the same high tone as an important lexical cue. The differing processing of tonal cues is reflected in a shorter MMN latency for the long vowel in the dialect group, illustrating higher sensitivity and accuracy in word discrimination due to the possibility to use the high tone in combination with the quantity cue for lexical retrieval.

Thus, we can show that a contrast which is phonologically relevant in two different varieties elicits similar electrophysiological effects in form of an MMN. However, this signature is further modulated by cues that are only relevant for one of the systems due to the regional linguistic background.

Acknowledgements: The research presented here was supported by the Hessian State Ministry of Higher Education, Research and the Arts, Landesoffensive zur Entwicklung wissenschaftlich-ökonomischer Excellenz (LOEWE), *Research Focus:*

12 Because of the overlong vowels and salient pitch movements due to Tone Accent 2, monosyllables might be perceived as bisyllabic words by non-native speakers, which entails an evaluation of the stressed syllable in the perceived word.

Exploring fundamental linguistic categories grant, research project *Phonological word – Constituents of the Phonological Word* granted to Joachim Herrgen, Jürgen Erich Schmidt, and Alexander Werth. The authors thank mayor Wolfgang Treis and the employees of the municipal administration in Mayen for their generous support in the field and Lydia Riedl and Corinna Schwarz for their help in preparing the stimuli.

References

Amenedo, Elena & Carles Escera. 2000. The accuracy of sound duration representation in the human brain determines the accuracy of behavioral perception. European Journal of Neuroscience 12. 2570–2574.

Bertoli, Sibylle, Jacek Smurzynski & Rudolf Probst. 2002. Temporal resolution in young and elderly subjects as measured by mismatch negativity and a psychoacoustic gap detection task. Clinical Neurophysiology 113. 396–406.

Boersma, Paul & David Weenink. 2012. PRAAT: Doing Phonetics by Computer (ver. 5.3.08) [Computer program]. Amsterdam: Institut for Phonetic Sciences.

Chandrasekaran, Bharath, Ananthanarayan Krishnan & Jackson T. Gandour. 2007. Mismatch negativity to pitch contours is influenced by language experience. Brain Research 1128. 148–156.

Chandrasekaran, Bharath, Ananthanarayan Krishnan & Jackson T. Gandour. 2009. Sensory Processing of Linguistic Pitch as Reflected by the Mismatch Negativity. Ear & Hearing 30(5). 552–558.

Cheour, Marie, Pirjo Korpilahti, Olga Martynova & A. Heikki Lang. 2001. Mismatch Negativity and Late Discriminative Negativity in Investigating Speech Perception and Learning in Children and Infants. Audiology & Neuro-Otology 6. 2–11.

Chládková, Kateřina, Paola Escudero & Silvia C. Lipski. 2013. Pre-attentive sensitivity to vowel duration reveals native phonology and predicts learning of second-language sounds. Brain and Language 126. 243–252.

Christmann, Corinna, Stefan Berti, Claudia Steinbrink & Thomas Lachmann. 2014. Differences in sensory processing of German vowels and physically matched non-speech sounds as revealed by the mismatch negativity (MMN) of the human event-related brain potential (ERP). Brain and Language 136. 8–18.

Clements, George N. & Samuel Jay Keyser. 1983. CV phonology: A Generative Theory of the Syllable. Cambridge: MIT Press.

Colin, Cécile, Ingrid Hoonhorst, Emily Markessis, Monique Radeau, Marianne de Tourtchaninoff, Aurélie Foucher, Gregory Collet & Paul Deltenre. 2009. Mismatch Negativity (MMN) evoked by sound duration contrasts: An unexpected major effect of deviance direction on amplitudes. Clinical Neurophysiology 120, 51–59.

Cooper, Rowena J., Juanita Todd, Katherine McGill & Patricia T. Michie. 2006. Auditory sensory memory and the aging brain: A mismatch negativity study. Neurobiology of Aging 27. 752–762.

Czigler, István & István Winkler. 1996. Preattentive auditory change detection relies on unitary sensory memory representations. NeuroReport 7. 2413–2417.

Dogil, Grzegorz. 1999. The phonetic manifestation of word stress in Lithuanian, Polish, German and Spanish. In Harry van der Hulst (ed.), Word prosodic systems in the languages of Europe, 273–311. Berlin & New York: Mouton de Gruyter.

Eulitz, Carsten & Aditi Lahiri. 2004. Neurobiological Evidence for Abstract Phonological Representations in the Mental Lexicon during Speech Recognition. Journal of Cognitive Neuroscience 16(4). 577–583.

Féry, Caroline. 2001. Phonologie des Deutschen: Eine optimalitätstheoretische Einführung, Teil II (Linguistics in Potsdam 11). Potsdam: Univ.-Bibliothek, Publ.-Stelle.

Fournier, Rachel, Carlos Gussenhoven, Ole Jensen & Peter Hagoort. 2010. Lateralization of tonal and intonational pitch processing: An MEG study. Brain Research 1328. 79–88.

Fuhrhop, Nanna & Jörg Peters. 2013. Einführung in die Phonologie und Graphematik. Stuttgart: Metzler.

Gandour, Jackson T. 2006. Tone: Neurophonetics. In Keith Brown (ed.), Encyclopedia of language and linguistics, vol. 12, 2nd edn., 751–760. Oxford: Elsevier.

Gandour, Jackson T. 2007. Neural substrates underlying the perception of linguistic prosody. In Carlos Gussenhoven & Tomas Riad (eds.), Tones and Tunes, Volume 2: Experimental Studies in Word and Sentence Prosody (Phonology and Phonetics 12.2), 3–25. Berlin & New York: Mouton de Gruyter.

Gandour, Jackson T., Yunxia Tong, Donald Wong, Thomas Talavage, Mario Dzemidzic, Yisheng Xu, Xiaojian Li & Mark Lowe. 2004. Hemispheric roles in the perception of speech prosody. NeuroImage 23. 344–357.

Gandour, Jack, Donald Wong, Li Hsieh, Bret Weinzapfel, Diana Van Lancker & Gary D. Hutchins. 2000. A Crosslinguistic PET Study of Tone Perception. Journal of Cognitive Neuroscience 12(1). 207–222.

Gussenhoven, Carlos. 2004. The phonology of tone and intonation (Research surveys in linguistics). Cambridge: Cambridge University Press.

Gussenhoven, Carlos & Jörg Peters. 2004. A tonal analysis of Cologne Schärfung. Phonology 21. 251–285.

Hayes, Bruce. 1999. Inalterability in CV Phonology. In John A. Goldsmith (ed.), Phonological Theory: The Essential Readings. Malden, MA & Oxford: Blackwell.

Heike, Georg. 1962. Suprasegmentale Merkmale der Stadtkölner Mundart. Phonetica 8. 147–165.

Huynh, Huynh & Leonard S. Feldt. 1976. Estimation of the Box Correction for Degrees of Freedom From Sample Data in Randomized Block and Split-Plot Designs. Journal of Educational Statistics 1. 69–82.

Hyman, Larry M. 2003. A Theory of Phonological Weight, 2nd edn. (The David Hume series). Stanford: CSLI Publications.

Jaramillo, Maria, Paavo Alku & Petri Paavilainen. 1999. An event-related potential (ERP) study of duration changes in speech and non-speech sounds. NeuroReport 10. 3301–3305.

Jaramillo, Maria, Titta Ilvonen, Teija Kujala, Paavo Alku, Mari Tervaniemi & Kimmo Alho. 2001. Are different kinds of acoustic features processed differently for speech and non-speech sounds? Cognitive Brain Research 12. 459–466.

Jessen, Michael, Krzysztof Marasek, Katrin Schneider & Kathrin Clahßen. 1995. Acoustic correlates of word stress. Proceedings of the 13th international congress of phonetic sciences 1995 (ICPhS), Stockholm. 428–431.

Kaan, Edith, Christopher M. Barkley, Mingzhen Bao & Ratree Wayland. 2008. Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: An event-related potentials training study. BMC Neuroscience 9. 53.
Kirmse, Ursula, Sari Ylinen, Mari Tervanieme, Martti Vainio, Erich Schröger & Thomas Jacobsen. 2008. Modulation of the mismatch negativity (MMN) to vowel duration changes in native speakers of Finnish and German as a result of language experience. International Journal of Psychophysiology 67. 131–143.
Kohler, Klaus J. 1979. Dimensions in the perception of fortis and lenis plosives. Phonetica 36. 332–343.
Köhnlein, Björn. 2011. Rule reversal revisited: synchrony and diachrony of tone and prosodic structure in the Franconian dialect of Arzbach. Utrecht: LOT.
Korpilahti, Pirjo, Christina M. Krause, Irma Holopainen & A. Heikki Lang. 2001. Early and Late Mismatch Negativity Elicited by Words and Speech-Like Stimuli in Children. Brain and Language 76. 332–339.
Korpilahti, Pirjo, A. Heikki Lang & Olli Aaltonen. 1995. Is there a late-latency mismatch negativity (MMN) component? Electroencephalography and Clinical Neurophysiology 95. P96.
Kutas, Marta & Kara D. Federmeier. 2011. Thirty years and counting: finding meaning of the N400 component of the event-related brain potential (ERP). Annu. Rev. Psychol. 62. 621–647.
Lass, Roger. 1984. Phonology: An introduction to basic concepts (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.
Lau, Ellen F., Colin Phillips & David Poeppel. 2008. A cortical network for semantics: (de)constructing the N400. Nature Reviews Neuroscience 9(12), 920–933.
Levänen, Sari, Riitta Hari, Linda McEvoy & Mikko Sams. 1993. Responses of the human auditory cortex to changes in one versus two stimulus features. Experimental Brain Research 97. 177–183.
Menning, Hans, Satoshi Imaizumi, Pienie Zwitserlood & Christo Pantev. 2002. Plasticity of the human auditory cortex induced by discrimination learning of non-native, moratimed contrasts of the Japanese language. Learning Memory 9(5). 253–267.
Näätänen Risto, Petri Paavilainen & Kalevi Reinikainen. 1989. Do event-related potentials to infrequent decrements in duration of auditory stimuli demonstrate a memory trace in man? Neuroscience Letters 107. 347–352.
Näätänen, Risto, Petri Paavilainen, Teemu Rinne & Kimmo Alho. 2007. The mismatch negativity (MMN) in basic research of central auditory processing: A review. Clinical Neurophysiology 118, 2544–2590.
Nenonen, Sari, Anna Shestakova, Minna Huotilainen & Risto Näätänen. 2003. Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing. Cognitive Brain Research 16. 492–495.
Nooteboom, Sieb. 1997. Prosody of speech: Melody and rhythm. In William J. Hardcastle & John Laver (eds.), The handbook of phonetic sciences, 640–673. Oxford: Blackwell.
Ohala, John J. 1972. How is pitch lowered? Journal of the Acoustical Society of America 52. 124 (A).
Ohala, John J. 1978. Production of Tone. In Victoria A. Fromkin (ed.), Tone: a linguistic survey, 5–39. New York et al.: Academic Press.
Ohala, John J. & William G. Ewan. 1973. Speed of pitch change. Journal of the Acoustical Society of America 53. 345 (A).

Partanen, Eino, Martti Vainio, Teija Kujala & Minna Huotilainen. 2011. Linguistic multifeature MMN paradigm for extensive recording of auditory discrimination profiles. Psychophysiology 48(10). 1372–1380.

Pekkonen, Eero, Teemu Rinne, Kalevi Reinikainen, Teija Kujala, Kimmo Alho & Risto Näätänen. 1996. Aging Effects on Auditory Processing: An Event-Related Potential Study. Experimental Aging Research: An International Journal Devoted to the Scientific Study of the Aging Process 22(2). 171–184.

Peters, Jörg. 2006. The Cologne word accent revisited. In Michiel de Vaan (ed.), Germanic Tone Accents (Zeitschrift für Dialektologie und Linguistik. Beihefte 131), 107–133. Stuttgart: Franz Steiner.

Phillips, Colin, Thomas Pellathy, Alec Marantz, Elron Yellin, Kenneth Wexler, David Poeppel, Martha McGinnis & Timothy Roberts. 2000. Auditory Cortex Accesses Phonological Categories: An MEG Mismatch Study. Journal of Cognitive Neuroscience 12(6). 1038–1055.

Pompino-Marschall, Bernd. 2003. Einführung in die Phonetik, 2nd edn. (de Gruyter Studienbuch). Berlin & New York: Walter de Gruyter.

Ramers, Karl-Heinz. 1998. Einführung in die Phonologie (UTB für Wissenschaft: Uni-Taschenbücher). München: Fink.

Ramers, Karl-Heinz & Heinz Vater. 1995. Einführung in die Phonologie (Kölner linguistische Arbeiten [KLAGE] – Germanistik 16). Hürth: Gabel.

Rimmele, Johanna, Elyse Sussman, Christian Keitel & Thomas Jacobsen. 2012. Electrophysiological Evidence for Age Effects on Sensory Memory Processing of Tonal Patterns. Psychology and Aging 27(2). 384–398.

Scharinger, Mathias, Philip J. Monahan & William J. Idsardi. 2011. You had me at "Hello": Rapid extraction of dialect information from spoken words. NeuroImage 56(4). 2329–2338.

Schiff, Sami, Pietro Valenti, Andrea Pellegrini, Maria Lot, Patrizia Bisiacchi, Angelo Gatta & Piero Amodio. 2008. The effect of aging on auditory components of event-related brain potentials. Clinical Neurophysiology 119. 1795–1802.

Schmidt, Jürgen Erich. 1986. Die mittelfränkischen Tonakzente (Mainzer Studien zur Sprach- und Volksforschung 8). Stuttgart: Franz Steiner.

Schmidt, Jürgen Erich. 2017. Neurodialektologie. Zeitschrift für Dialektologie und Linguistik 83(1). 56–91.

Schröger, Erich. 1998. Measurement and interpretation of the mismatch negativity. Behavior Research Methods, Instruments, & Computers 30(1). 131–145.

Schröger, Erich, Iria SanMiguel & Alexandra Bendixen. 2013. Prädiktive Modellierung in der auditiven Wahrnehmung. In Erich Schröger & Stefan Koelsch (eds.), Affektive und Kognitive Neurowissenschaft (Enzyklopädie der Psychologie 5), 11–45. Göttingen et al.: Hogrefe Verlag für Psychologie.

Sittiprapaporn, Wichian. 2012. Source Localization of Preattentive Processing for Different Vowel Duration Changes with Contour Tones in Monosyllabic Thai Words. Journal of Applied Sciences 12(15). 1580–1587.

Stock, Eberhard & Christina Zacharias. 1982. Deutsche Satzintonation, 3rd edn. Leipzig: VEB Verlag Enzyklopädie.

Sundberg, Johan. 1973. Data on maximum speed of pitch changes. Speech Transmission Laboratory – quarterly progress and status report 14. 39–47.

Sundberg, Johan. 1979. Maximum speed of pitch changes in singers and untrained subjects. Journal of Phonetics 7. 71–79.

Takegata, Rika, Mari Tervaniemi, Paavo Alku, Sari Ylinen & Risto Näätänen. 2008. Parameter-specific modulation of the mismatch negativity to duration decrement and increment: Evidence for asymmetric processes. Clinical Neurophysiology 119. 1515–1523.

Trubetzkoy, Nikolaj S. 1939. Grundzüge der Phonologie. Prag.

Verleger, Rolf, Winfried Neukäter, Detlef Kömpf & Peter Vieregge. 1991. On the reasons for the delay of P3 latency in healthy elderly subjects. Electroencephalography and clinical Neurophysiology 79. 488–502.

Werth, Alexander. 2011. Perzeptionsphonologische Grundlagen der Prosodie: Eine Analyse der mittelfränkischen Tonakzentdistinktion (Zeitschrift für Dialektologie und Linguistik. Beihefte 143). Stuttgart: Franz Steiner.

Werth, Alexander. 2012. Perceptual evidence for contrast enhancement in tone-intonation interaction. In Oliver Niebuhr (ed.), Understanding prosody, 187–200. Berlin & New York: Walter de Gruyter.

Wiese, Richard. 1996. The Phonology of German. Oxford: Clarendon Press.

Wiese, Richard. 2011. Phonetik und Phonologie (LIBAC. Linguistik für Bachelor). Paderborn: Fink.

Wolff, Christian & Erich Schröger. 2001. Human pre-attentive auditory change-detection with single, double, and triple deviations as revealed by mismatch negativity additivity. Neuroscience Letters 311(1). 37–40.

Ylinen, Sari, Minna Huotilainen & Risto Näätänen. 2005. Phoneme quality and quantity are processed independently in the human brain. NeuroReport 16. 1857–1860.

Ylinen, Sari, Anna Shestakova, Paavo Alku & Minna Huotilainen. 2005. The perception of phonological quantity based on durational cues by native speakers, second-language users and nonspeakers of Finnish. Language and Speech 48(3). 313–338.

Ylinen, Sari, Anna Shestakova, Minna Huotilainen, Paavo Alku & Risto Näätänen. 2006. Mismatch negativity (MMN) elicited by changes in phoneme length: A cross-linguistic study. Brain Research 1072. 175–185.

Zec, Draga. 1994. Sonority constraints on prosodic structure. New York & London: Garland.

Zhang, Jie. 2002. The effects of duration and sonority on contour tone distribution: A typological survey and formal analysis (Outstanding dissertations in linguistics). New York & London: Routledge.

Appendix

Tab. 5: Peak latencies and amplitudes of the late MMN at Fz for all experimental conditions

deviant$_{\text{/standard}}$	**group**	**peak latency at Fz**	**amplitude at Fz deviant-standard**
[ʃa̱ː^{2}l]$_{\text{/ʃa̱lˑ2/}}$	Dialect	430 ms	−0.535 µV
[ʃa̱ː^{2}l]$_{\text{/ʃa̱lˑ2/}}$	Standard German	356 ms	−0.780 µV
[ʃa̱lˑ^{2}]$_{\text{/ʃa̱ː2l/}}$	Dialect	554 ms	−1.162 µV
[ʃa̱lˑ^{2}]$_{\text{/ʃa̱ː2l/}}$	Standard German	612 ms	−1.307 µV

Natalie Boll-Avetisyan

The role of phonological structure in speech segmentation by infants and adults: a review and methodological considerations

Abstract: Across languages, phonological words are restricted by constraints on their phonological structure. As a by-product, word-internal phonological structure creates cues to word boundaries in continuous speech: if an acoustic sequence occurs in speech that is illegal or unlikely within words, then it is most likely that it resides at a word boundary. There is experimental evidence that infant and adult listeners exploit knowledge of phonological structure for segmenting speech into words. Results of such studies can, conversely, inform us about the nature of the phonological word, the relevant underlying mental representations, and when these are acquired. The present chapter consists of two parts. First, a literature overview of studies is presented that provide empirical evidence for a role of phonological structure in speech segmentation by infants and adults. The second part focuses on how effects of phonological structure on speech segmentation can be tested experimentally. An overview of methods is provided that can be used with infant and adult populations (i.e., artificial language segmentation, word spotting and methods using more naturalistic spoken text passages), and their pros and cons and their limits are evaluated. The chapter ends with methodological recommendations.

Keywords: speech segmentation, artificial language learning, phonological structure, infants, language acquisition

1 The speech segmentation problem and the role of phonological structure

When listening to our native language, it escapes us that the acoustic speech signal we perceive is actually continuous and does not contain clear-cut cues for word boundaries (e.g., Cole and Jakimik 1980; Goldsmith 1976; Harris 1944;

Natalie Boll-Avetisyan: Department Linguistik, Universität Potsdam, Karl-Liebknecht-Straße 24–25, D-14476 Potsdam, nboll@uni-potsdam.de

DOI 10.1515/9783110542899-009

Liberman and Prince 1977). Nonetheless, it is obvious to us that utterances are made up of individual words. Orthographic systems reflect our knowledge about where a word starts and ends by using spaces to mark word boundaries. The observation that speech is continuous and needs to be segmented in order to prepare lexical access has led to one of the most challenging and interesting research questions in the field of psycholinguistics: the *speech segmentation problem* (e.g., Cutler and Mehler 1993; Cutler 1994).

The speech segmentation problem may be most challenging for language-learning infants. While adults can rely on feedback from word knowledge when processing the speech signal (e.g., Marslen-Wilson and Welsh 1978; Mattys, White, and Melhorn 2005), the question arises of how language-learning infants lacking in lexical knowledge would segment the speech signal. If speech segmentation were only possible after lexical recognition, then pre-lexical infants, lacking lexical knowledge, would face an irresolvable task. However, instead, the capacity to segment speech must be a pre-requisite for lexical acquisition for pre-lexical infants. Speech segmentation is also a challenge for learners of a second language (L2). Anyone who has had foreign language training may share the experience that understanding foreign speech uttered by native speakers is anything but easy, especially in the initial phases of L2 learning. This problem may in part relate to the relatively small L2 lexicon, but also to the fact that segmentation routines acquired for the L1 need to be suppressed when listening to an L2.

Over the past decades, a number of studies have dealt with the speech segmentation problem. By means of this research, it has become evident that phonological structure offers important cues for speech segmentation. That is, listeners' knowledge of constraints on the internal phonological organization of words gives them clues about where to expect word endings and word beginnings in speech. Experimental and theoretical phonologists often do not consider that knowledge of the structure of phonological words may be put to use in this way in speech and language processing. Moreover, most of the empirical work on speech segmentation is carried out by psycholinguists, who might not necessarily consider that their findings may be informative about the type of underlying phonological representations that influence the detection of word boundaries in speech. The aim of this chapter is to raise attention to the potential of studying effects of underlying knowledge of phonological words empirically by testing its effects on speech segmentation.

The aim of the current chapter is two-fold: first, it aims at providing an extensive overview of the studies that have focused on the speech segmentation problem and provide empirical evidence for the role of phonological structure in speech segmentation (Section 2). Second, it aims at providing the reader with an overview of the experimental methods (*artificial language* [henceforth *AL*] learning paradigm, *word spotting* paradigm) that can be applied for testing the effects of phonological structure on speech segmentation (Section 3). Previous review articles have already discussed how effects of phonological structure on language *learning* can be studied using the AL learning paradigm (Moreton and Pater 2012). Furthermore, some 20 years ago, McQueen has written a review article to introduce the word spotting paradigm as a method (McQueen 1996). To my knowledge, there is, as yet, no paper that has attempted to offer a comprehensive review of studies that are informative of the role of phonological structure in *speech segmentation* and that has compared the different experimental methods that can be used for testing the effects of phonological structure on speech segmentation empirically. It is the aim of the current chapter to fill this gap.

The chapter is structured as follows: in Section 2, It will be discussed how word-internal phonological structure provides cues for speech segmentation (Section 2.1). Next, in Section 2.2, literature covering the use of phonological cues for segmentation by adults and infants will be reviewed. Thereby I will also discuss the interplay of potentially innate assumptions about the phonological structure of phonological words and their interaction with language-specificity. The second part of this chapter (Section 3) discusses which experimental methods can be used for studying research questions regarding the use of phonological structure for speech segmentation by adults and infants. First (in Section 3.1), an overview of the methods is provided. After this, methodological advantages of the different segmentation tasks are discussed, and it is explained why the AL segmentation paradigm is an ideal method for studying which underlying phonological representations of the structure of phonological words guide our speech segmentation. This last section ends with offering recommendations with regards to designing AL learning experiments. The chapter ends with a conclusion (Section 4).

The chapter should be of interest for phonologists, psycholinguists and developmental psychologists. It should be of particular use for students and researchers who are planning their own studies on effects of phonological or phonetic knowledge on language acquisition and speech processing.

2 Word-internal phonological structure and speech segmentation

2.1 The phonological word, its demarcative function and role in language acquisition

Formal phonology has proposed that there must be a unit larger than the syllable or the metrical foot, but smaller than the phonological phrase. This unit is the phonological word (also referred to as prosodic word or pword; onwards this will be referred to as the phonological word or simply as the word). The theoretical and formal arguments that justify the assumption of the phonological word are that there are specific characteristics that apply to the domain of phonological words, namely word stress rules, segmental word-level rules and phonotactics (Hall 1999; Nespor and Vogel 1986; Peperkamp 1999). The boundaries of phonological words mostly coincide with that of lexical words, and these phonological rules and generalizations do not only apply to monomorphemic words, but often span over the whole word independent of its morphological complexity.

As discussed in other chapters of this volume, knowledge of phonological constraints on phonological words affects lexical processing both in production and comprehension (see the chapters by Bergmann 2018, by Domahs, Domahs, and Kauschke 2018, and by Scharinger 2018 in this volume). Moreover, it affects lexical learning (see, e.g., Ulbrich and Wiese 2018 in this volume). Furthermore, as we will see in the present chapter, it also affects lexical segmentation. The fact that phonological words are characterized by rules, generalizations and constraints on their phonological structure has an impact on the phonological structure of speech that, in consequence, contains evident cues to word boundaries. For the illustration, consider these examples:

(1) Prosody
Some languages restrict their lexicon by fixed lexical stress. Finnish, for example, requires that words start with a stressed syllable. Hence, if the word *HYmy* is being preceded by a weak syllable (e.g., *puHYmy*), then evidently the syllable /hy/ must mark the onset of a new word.

(2) Phonotactics
Some languages restrict their lexicon by vowel harmony. Finnish, for example, requires that vowels within words are either all front vowels or all back vowels. Hence, if the word *hymy* is being followed by a syllable containing a back vowel

(e.g. *hymypu*), then evidently the syllable /pu/ must be part of a new word (example taken from Suomi, McQueen, and Cutler 1997).

The observation that phonological structure may have a demarcative function is not new. Already Trubetzkoy (1939/1969) mentioned that phonological constraints like vowel harmony or fixed lexical stress could be useful for listeners when segmenting speech, and, long before experimental evidence supported this claim, it was speculated that such word-internal phonological restrictions may contribute to language acquisition by helping children to discover linguistic subunits in the input (Peters 1985; Trubetzkoy 1939/1969). In the 1980s and 1990s, these ideas developed into the *phonological bootstrapping account*, that is, the concrete proposal that infants make use of phonological knowledge to initiate lexical and syntactic acquisition (Gleitman and Wanner 1982; Morgan and Demuth 1996; Weissenborn and Höhle 2001).

The bootstrapping account is supported by much of what we know about early language development. From birth, infants are sensitive to phonological properties of language like phonetic contrasts (Lasky, Syrdal-Lasky, and Klein 1975) and rhythmic patterns (e.g., Nazzi, Bertoncini, and Mehler 1998). During the first half of the first year of life, infants start acquiring the sound structure of their ambient language, including the acquisition of prosodic knowledge, the formation of phonemic categories and the acquisition of language-specific phonotactics (for an overview, see Jusczyk 1997). Hence, by the age of 6 months, infants have the pre-requisites for using their knowledge of phonological structure as a bootstrap for lexical and syntactic acquisition.

In some phonological frameworks it has been suggested that phonological rules and constraints are derived from the lexicon (e.g., Hayes and Wilson 2008, Pierrehumbert 2003). If that were the case, then infants would only benefit from facilitatory cues to segmenting speech into proto-words after lexical acquisition. However, in principle, knowledge of the phonological structure of words could also be directly induced from the continuous speech stream (e.g., Adriaans and Kager 2010; Brent and Cartwright 1996; Cairns et al. 1997; Perruchet and Vinter 1998; Swingley 2005).

The speech segmentation problem also arises when adults learn an L2. Adults acquire L2s in different contexts, and this context may also have different effects on the segmentation problem. If adults learn an L2 by means of explicit instructions in a class-room setting, they may have one advantage: their initiation of L2 lexical acquisition does not depend on the ability to segment speech of the target language. Particularly in systematic formal language education, adult learners are taught lists of isolated words. Lexical knowledge acquired in such a way may be helpful when segmenting L2 speech. However, if adults acquire an L2 in an immersion context, for example, once they move to the country where the target language is spoken, the L2 is acquired under conditions that are more similar to L1 acquisition. The issue of speech segmentation in an L2 has not yet received much attention in the literature, so little is known about how L2 speech segmentation abilities are acquired. One study that addressed this issue found that learners who acquire their L2 in an immersion context are better at acquiring relevant sublexical cues for speech segmentation than learners who learn the L2 under the condition of formal instruction (Boll-Avetisyan 2012). There is, moreover, some evidence that the amount of current daily exposure to an L2 as well as L2 acquisition contexts in which learners receive much auditory input is associated with a more target-language-like rhythmic perception of speech (Boll-Avetisyan et al. 2016).

A question that will be of interest for phonologists is how potentially innate assumptions about the phonological structure of phonological words interact with language-specific knowledge. Most phonologists assume that typologically recurrent phonological structures reflect a cognitive endowment for having an advantage for acquiring these structures. Some experimental studies have provided evidence for influences of innate biases on acquisition in infants and adults (e.g., Abboub et al. 2016; Bion, Benavides-Varela, and Nespor 2011; Frost, Monaghan, and Tatsumi 2017; Moreton 2008; Ulbrich et al. 2016; White and Sundara 2014; Wiese et al. 2017). It seems that these biases reduce or lose their effect on speech perception if not supported by the properties of the native language (e.g., Bhatara et al. 2013; Boll-Avetisyan and Kager 2014), but their influence can, as it seems, be reactivated when learning an AL or an L2 (e.g., Boll-Avetisyan et al. 2016; Ettlinger, Finn, and Hudson Kam 2012; Moreton 2008; and Ulbrich and Wiese, this volume). The question arises of how innate biases and language-specific phonological knowledge may interact in helping us to distinguish between relevant and irrelevant information in continuous speech, thereby reducing the hypothesis space about which stretches of sound in speech might be phonological words. Only few studies have directly dealt with these issues. Studies with bilinguals, whose one language is constrained by a specific naturally preferred structure, while the other is not, can give interesting insights to this question (Boll-Avetisyan 2012; Cutler et al. 1989; Cutler et al. 1992).

In the following, experimental literature will be reviewed that shows that the word-internal phonological organization plays an important role for adults and infants when identifying words in continuous speech. The reader should keep in mind that these studies are not only indicative of human processing abilities.

They also inform us how phonology is mentally represented and about whether phonologists' assumptions of phonological structure have a psychological reality.

2.2 Experimental evidence for the use of phonological cues for segmentation

Since the 1990s, a growing body of psycholinguistic studies has gathered evidence that sub-lexical knowledge about word-internal phonological structure can be used as a cue for speech segmentation. In the following, we will review studies that have assessed the use of prosodic and phonotactic cues for speech segmentation by adults and infants. Two types of experimental tasks are typically used in order to test adults' use of sub-lexical cues for speech segmentation: the word spotting task and the AL segmentation task. With infants, AL segmentation tasks can also be used. An alternative is a task with more naturalistic stimuli which combines a use of natural stretches of text and isolated words as stimuli. These methods will be described in detail in Section 3.

2.2.1 Prosodic cues for segmentation

The first studies that assessed whether listeners exploit sublexical cues to segment words from speech investigated the role of prosodic cues. The pioneer study by Cutler and Norris (1988) investigated whether listeners make use of metrical cues in the speech signal. As the large majority of English words start with a strong syllable, it was hypothesized that native listeners of English assume the beginning of a word when hearing a strong syllable. In a word spotting task, participants were slower in finding words in disyllables in which both syllables were strong (e.g., the word *mint* in *mintayve*) than when just the first syllable was strong (e.g., *mint* in *mintesh*). The authors interpret the result in support of their hypothesis: when hearing strong syllables in speech, listeners will attempt lexical access. When immediately hearing a second strong syllable, this second lexical access attempt will inhibit the activation of the precedingly heard *mint*, hence the slower reaction time. Similar results are found in experiments with native listeners of other stress-timed languages, in which words are typically trochaic like Dutch (e.g., Vroomen, van Zon, and de Gelder 1996), while native speakers of syllable-timed languages like French apply a syllable-based segmentation strategy (Mehler et al. 1981). There is some evidence from experiments by Cutler and colleagues (Cutler et al. 1989; Cutler et al. 1992) with simultaneous French-English bilingual adults that the metrical segmentation strategy is actually the unmarked (i.e., universally preferred) segmentation procedure. Only French-dominant bilinguals apply the syllable-based segmentation strategy, and they only do so, when the speech material is French, while they apply the metrical segmentation strategy when listening to English material. English-dominant bilinguals, however, always apply the metrical segmentation strategy independent of whether they hear French or English.

Cutler and colleagues (e.g., Cutler and Norris 1988; Cutler and Butterfield 1992; Norris, McQueen, and Cutler 1995) suggested that this metrical segmentation cue mainly results from listeners attending to full versus reduced schwa vowels in speech. Later research has indicated that other prosodic cues such as pitch accent and final lengthening are also used for finding words in speech. This research has often implemented the AL learning paradigm, as it is relatively easy to manipulate the relative prosodic cues in artificial speech, and they have often employed cross-linguistic settings by including in their test population speakers of languages differing in prosody. One study found that word-initial pitch accent is used to identify word onsets by Dutch- and Finnish-speaking adults, but not by native speakers of French, which is predicted by their native languages, as Dutch is trochaic, Finnish has fixed initial stress, but French has no lexical stress (Vroomen, Tuomainen, and de Gelder 1998). Another study (Tyler and Cutler 2009) compared English, French and Dutch speakers on their use of pitch accent and lengthening cues for segmentation. They found that all three groups used word-final lengthening cues, but no group used initial lengthening cues for speech segmentation, which corresponds to the final lengthening cue being universally used for marking final boundaries. With regards to pitch, cross-linguistic differences were found: English listeners only used word-initial pitch cues for segmentation, French listeners only used word-final pitch cues, and Dutch listeners used both.

One recent study (Tremblay et al. 2016) has, for the first time, investigated the use of knowledge of L2 lexical prosody for speech segmentation. In this study, it was found that native English L2 learners of French were better at using final pitch rise as a cue for segmenting French than Korean L2 learners of French. The results may seem counterintuitive, as Korean, like French, has phrase-final pitch rise, while English listeners' should have a preference for initial pitch rise from their L1. The authors suggest that it may actually be easier to acquire L2 stress cues for segmentation if L1 stress cues are very different. More research in this field will be needed to verify these findings.

Infants are sensitive to prosody from birth (Mehler et al. 1988; Nazzi, Bertoncini, and Mehler 1998), and they establish language-specific prosodic knowledge from 6 months onwards (Höhle et al. 2009). Much evidence suggests that a language-specific use of prosodic cues for segmentation develops between 7 and 10 months. Infants, whose ambient language is stress-timed with words typically being trochaic, such as American and Canadian English and Dutch (Ramus, Nespor, and Mehler 1999), can segment words with a strong-weak (trochaic) stress pattern from speech (Houston et al. 2000; Jusczyk, Houston, and Newsome 1999; Kooijman, Hagoort, and Cutler 2009). They can also segment trochaic words from non-native speech (Houston et al. 2000), even if this non-native language is rhythm-timed like Italian (Pelucchi, Hay, and Saffran 2009). At 6 months they can segment trochees if the words coincide with utterance boundaries (Johnson, Seidl, and Tyler 2014). A recent study with 9 month old German-learning infants, however, suggests that lexical stress by itself is not a sufficient cue for segmentation, but that the stressed syllable should additionally be marked by pitch accent (Zahner, Schönhuber, and Braun 2016).

Dutch- and English-learning infants' ability to segment words with a weak–strong (iambic) pattern develops somewhat later, beginning at the age of 10 months (Jusczyk, Houston, and Newsome 1999; Kooijman, Hagoort, and Cutler 2009), although a recent study found that they can segment iambic words already at 7.5 months in the context of additional cues from the lexicon, that is, if the iamb is preceded by a familiar word such as *mommy* (Sandoval and Gómez 2016). It seems straightforward that these differences in infants' readiness to segment trochees versus iambs may suggest that speech segmentation is initially guided by the typical prosodic patterns of the ambient language, as iambs are less frequent than trochees in these languages.

More evidence suggesting that the use of prosodic cues for segmentation in infancy is language-specific comes from studies with French-learning infants. French is a rhythm-timed language without lexical stress (Féry, Hörnig, and Pahaut 2011; Ramus, Nespor, and Mehler 1999). When familiarized with isolated words and tested on text passages, French-learning infants did not segment disyllables from speech until the age of 16 months. At earlier ages, they only segment monosyllables, which suggests that they recognize the syllable as the basic rhythmic unit (Nazzi et al. 2006). Follow-up work, however, showed that French-learning infants can segment disyllabic words at an earlier age when tested under specific conditions. ERP paradigms seem to be more sensitive to revealing infants' recognition of disyllabic words at 12 months (Goyet, de Schonen, and Nazzi 2010). Moreover, when familiarized with passages and tested on their recognition of isolated disyllabic words (that is, reversing the order which was used by Nazzi et al. 2006), already 8-month-olds succeed (Nazzi et al. 2014). Such results highlight the importance of comparing different methods in hypothesis-testing.

There is, furthermore, evidence for infants learning an iambic language to segment iambs from speech. As opposed to European French, Canadian French is considered to have fixed iambic lexical stress. Polka and Sundara (2012) compared Canadian French- and Canadian English-learning infants at 8 months of age, who were familiarized with isolated disyllabic words and tested on text passages (as in Nazzi et al. 2006). Both groups of infants succeeded in segmenting disyllables when tested on a stimulus set recorded in their native language. However, neither of them recognized the familiarized words in speech when they were tested in their non-native language. Since the disyllables were iambs in the Canadian French stimulus set, but trochaic in the Canadian English stimulus set, these results can be interpreted as evidence for Canadian French-learning infants to have the ability to segment iambs but not trochees from speech. Although the study is the first that may constitute evidence for an iambic segmentation strategy by infants learning an iambic language, the authors themselves do not discuss their results in this way. They suggest that differences between European and Canadian French-learning infants may relate to slight differences in the experimental procedure between Nazzi et al.'s (2006) and their study.

While the above-mentioned studies have used naturalistic segmentation tasks, other studies have found comparable results using AL segmentation paradigms (Curtin, Mintz, and Christiansen 2005). AL studies have also directly compared the use of statistical and prosodic cues for speech segmentation. Results suggest that young 7-month-old infants prefer relying on statistical cues (Thiessen and Saffran 2003), but at 8 and 9 months, they prefer relying on prosodic cues (Johnson and Jusczyk 2001; Thiessen and Saffran 2003). The authors take their results to suggest that the ability to use prosody for segmentation is developed later than the ability to discover distributional cues for segmentation in speech.

The above mentioned studies suggest a development of a use of prosodic cues for speech segmentation in infancy. Other studies indicate that infants may, however, start out segmenting speech by relying on a universal bias, the Iambic/Trochaic Law (Hayes 1995; Nespor et al. 2008), a principle guiding the grouping of syllables into larger units based on their duration, pitch, or intensity patterns. This work has shown that infants use these rhythmic cues to segment words from ALs, and there is, as yet, no clear evidence for cross-linguistic differences (Abboub et al. 2016; Bion, Benavides-Varela, and Nespor 2011; Hay and Saffran 2012). Bion, Benavides-Varela, and Nespor (2011) exposed Italian-learning 7.5-month-olds to a continuous stream of syllables alternating in either pitch or duration. At test, infants showed a preference for disyllables that had been presented with high-low pitch patterns, but they did not use duration (Bion, Benavides-Varela, and Nespor 2011). In a similar procedure, both French- and German-learning 7.5-month-olds indicated segmentation preferences for disyllables of a high-low pitch pattern and, under certain circumstances (that were not part of Bion, Benavides-Varela, and Nespor's experimental design), a short-long duration pattern, but they did not use intensity (Abboub et al. 2016). Hay and Saffran (2012) have explored infants' responses to intensity and duration variation when statistical cues are present and showed that 9-month-old infants use both cues, and that 6.5-month-olds use intensity variation.

2.2.2 Phonotactic cues for segmentation

A large number of studies suggest that listeners make use of different types of phonotactic restrictions to detect word boundaries in speech. Some studies have, for example, investigated the use of vowel harmony as a segmentation cue. Using word spotting, Suomi, McQueen, and Cutler (1997) showed that adult Finnish listeners were faster in detecting words (e.g., *hymy*) in speech if preceded by a syllable containing a non-harmonic vowel (e.g. in *puhymy*) than if preceded by a syllable containing a harmonic vowel (e.g. in *pyhymy*). Vroomen, Tuomainen, and de Gelder (1998) first replicated Suomi, McQueen, and Cutler's (1997) findings. Next, they compared Finnish, French and Dutch listeners in an AL learning experiment, in which vowels were harmonic within words in one condition (creating disharmonic transitions between words), but disharmonic within words in the other condition. As a result, only Finnish but not French and Dutch listeners enhanced their segmentation performance when the AL contained cues from vowel harmony. These findings suggest that the use of vowel harmony must be acquired from the language, as Finnish, but neither Dutch nor French restricts its lexicon by vowel harmony. Comparable results were obtained in cross-linguistic study using a nonword spotting task: here only Turkish (a language with vowel harmony) but not French listeners made use of cues from vowel harmony (Kabak, Maniwa, and Kazanina 2010).

Other studies suggest the relevance of position-specific restrictions on vowels in speech segmentation. For example, Italian listeners were slower in spotting vowel-final words if they occurred before a carrier that started in a short vowel [u] that had to be interpreted as part of a diphthong with the preceding vowel (e.g., the word *città* in *cittàu.ba*) than when it occurred before a carrier that started in a short [o] (e.g., *città* in *città.oba*), creating a diphthong illicit in Italian (Tagliapietra et al. 2009). Moreover, English listeners detect word boundaries by relying on knowledge of the cue that English words cannot end in lax vowels (Skoruppa et al. 2015).

Consonant phonotactic knowledge also serves as a cue for speech segmentation. The seminal study by McQueen (1998) showed that listeners exploit phonotactic information about specific adjacent consonants when segmenting speech. Dutch listeners were faster in detecting words like *rok* when embedded in a phonotactic context resulting in an onset or coda consonant cluster illegal in Dutch (e.g., /mr/ in *fimrok*) than when embedded in a legal context (e.g., /dr/ *fidrok*). These results were replicated in a study that added eye-tracking as an additional measure (Lentz 2011). Similar results were obtained with Italian listeners (Tagliapietra et al. 2009). Moreover, one study showed that Korean listeners have a graded sensitivity to segmentation cues depending on whether a cluster would be likely or unlikely to contain a word boundary or whether it would be even illicit to insert a boundary (Warner et al. 2005).

Other studies have used the AL segmentation paradigm for testing effects of consonant phonotactics on segmentation. Two studies directly compared phonotactic and statistical cues of word boundaries (Ettlinger, Finn, and Hudson Kam 2012; Finn and Hudson Kam 2008). Finn and Hudson Kam (2008) created an AL, in which words started with onset clusters that were either licit or illicit in English. Results suggested that English listeners assumed boundaries between consonants in illicit clusters in spite of the presence of transitional probability cues for keeping them intact. In their follow-up study (Ettlinger, Finn, and Hudson Kam 2012), they tested the effects of the Sonority Sequencing Principle (Selkirk 1982) on segmentation. All words had onset clusters that were illicit in English, but some violated the Sonority Sequencing Principle while others did not. Results suggested that participants can segment words that do not violate the Sonority Sequencing Principle, keeping the universally preferred consonant clusters intact. These studies indicate that listeners may transfer both language-specific and universal phonotactic knowledge about consonant clusters when listening to an unknown language speech stream.

Two further studies investigated the use of phonotactic constraints on non-adjacent consonants for speech segmentation (Boll-Avetisyan and Kager 2014; Boll-Avetisyan and Kager 2016). One showed that listeners of Dutch insert word boundaries between non-adjacent labials (e.g., /bVm/, /pVb/ etc.) in an AL stream, which is expected as co-occurrences of non-adjacent labial consonants are improbable within words in Dutch (Boll-Avetisyan and Kager 2014). Their follow-up study (Boll-Avetisyan and Kager 2016) assessed whether these results reflected a use of an abstract phonotactic constraint capturing all labial-labial-co-occurrences (i.e., OCP-LABIAL) or whether segmentation preferences rather reflect knowledge of occurrence probability of specific consonant pairs. Here, experiments involved specific labial-vowel-labial pairs that were either over- or underattested in Dutch. Results revealed that Dutch listeners did not insert boundaries between labials that were highly likely to co-occur in Dutch, suggesting a use of specific rather than abstract knowledge. However, their reanalysis of Boll-Avetisyan and Kager's (2014) data (adding specific co-occurrence probabilities as a covariate) did not show an effect of specific probabilistic knowledge on segmentation in the original study. The authors suggest that phonotactic knowledge is represented as both abstract constraints (e.g., OCP-LABIAL) and specific distributional knowledge (e.g., high likelihood of /pVp/, low likelihood of /pVm/), that both types of knowledge can affect speech segmentation, and that it depends on processing demands whether listeners draw on abstract or specific knowledge.

Few studies have assessed the effect of knowledge of an L2 on segmentation. One study (Weber and Cutler 2006) looked at the use of phonotactic cues from specific consonant clusters indicating boundaries in either just the L1 or just the L2. Results showed that German-speaking proficient learners of English use both German and English phonotactic cues for detecting English words in a word spotting task. Similar results were obtained in a study with Dutch L2 learners of English (Lentz 2011). This indicates that L2 learner can acquire L2 segmentation strategies, but they fail at suppressing a reliance of their L1 phonotactic cues. Boll-Avetisyan (2012) reports that Mandarin Chinese is not constrained by OCP-LABIAL, and accordingly Mandarin Chinese listeners do not use the constraint for segmentation. However, they use OCP-LABIAL for segmentation if they acquired Dutch as an L2, but only under the condition that they acquired Dutch while living in the Netherlands and not when they learned Dutch at the university while living in China.

Several studies provide evidence for infants' use of phonotactic cues for speech segmentation. Some studies have, for example, shown that infants at first only segment consonant- but not vowel-initial words. English-learning infants cannot segment vowel-initial words from speech until 13.5 or 16 months of age (Mattys and Jusczyk 2001b; Nazzi et al. 2005), unless they occur in salient phrase-initial or phrase-final positions (Seidl and Johnson 2008). French-learning infants even only succeed at 24 months (Babineau and Shi 2014). German-learning infants, on the contrary, already segment both vowel- and consonant-initial words at an age of 11 months (Boll-Avetisyan, Fritzsche, and Jäkel 2017). These early segmentation preferences may reflect distributional knowledge of the native language, as vowel-initial words are more common in German than in English, and word beginnings are particularly intransparent in French because of liaison. However, preferences for consonant-initial segmentations may also suggest an influence of an innate bias during early infancy, as typologically, across languages, more words start with consonants, and all languages that have vowel-initial words also have consonant-initial words (Blevins 1995).

Other studies investigated the use of consonant cluster probability for segmentation. Mattys et al. (1999) exposed infants to CVCCVC nonwords with internal clusters of high versus low probability. Results suggested that infants preferred those items with clusters of a high phonotactic probability. When, however, a boundary (pause of 500 ms) was inserted between the two consonants (i.e., CVC__CVC), the effect was reversed: now infants preferred listening to items, in which there was a clear boundary between consonants that are unlikely to occur within words. In a second study (Mattys and Jusczyk 2001a), they familiarized 9-month-olds with text passages with nonwords embedded in a context of consonants that would either result in a cluster with high or low probability of occurrence. Afterwards, they were exposed to isolated target and control items, and infants showed a preference for listening to target items that had occurred in the context of a phonotactic segmentation cue. These two studies suggest that infants use knowledge of adjacent consonant co-occurrence probabilities for speech segmentation.

Other studies explored the use of knowledge of non-adjacent dependencies for segmentation. One study showed that Turkish- but not German-learning 9-month-olds use vowel harmony as a segmentation cue, which corresponds to the fact that the Turkish but not the German lexicon is constrained by vowel harmony (Van Kampen et al. 2008). Moreover, at 10 months, infants begin using knowledge of non-adjacent consonant co-occurrence probabilities. French-learning infants are better at identifying words starting with a labial-vowel-coronal pattern (which is highly frequent in French) than if they start with a coronal-vowel-labial pattern (which are less likely in French; Gonzalez-Gomez and Nazzi 2013).

While all of the above mentioned infant studies have used naturalistic stimuli to explore effects of phonotactics in infants' segmentation, one study has used an AL paradigm, and found that Dutch-learning 15-month-olds are more likely to insert boundaries between non-adjacent consonants with a low co-occurrence probability than that they use cues from a high co-occurrence probability for keeping the sequence as a unit. This result suggests that their preferred segmentation strategy is to find the boundaries in speech rather than finding the chunks (Boll-Avetisyan 2012). Moreover, one study investigated whether Dutch-learning infants use OCP-Labial for segmenting an AL, but they failed to find evidence for this (Boll-Avetisyan et al. 2009). Evidently, the failure may be a consequence of methodological issues. However, the authors also point out that the Dutch-directed lexicon contains many high-frequency words in which labials are reduplicated (*papa* 'dad', *mama* 'mom', *pap* 'porridge', *popje* 'doll'), and they demonstrate by a corpus analysis that the effect of OCP-LABIAL is much weaker in infant-directed than in adult-directed speech.

2.3 Synthesis

It was discussed that the ability to segment speech by relying on sublexical cues may be crucial for lexical and syntactic acquisition. It was illustrated how properties of phonological words give sublexical cues to word boundaries in speech. Next, experimental studies were reviewed that have dealt with infants' and adults' ability to use cues from prosody and phonotactics for finding word boundaries in speech. While most of these studies indicate a use of language-specific knowledge of word-internal phonological structure, some also suggested an influence of innate biases. The next part of the chapter is targeted at evaluating the experimental methods that are used for studying effects of phonological structure on word segmentation.

3 Psycholinguistic methods for studying speech segmentation

3.1 Artificial language segmentation

There are ALs, such as Esperanto or Klingon, which are constructed, human-made languages that people actually learn and use for communication. The kind of ALs that psycholinguists use for studying speech segmentation have very little to do with them. ALs constructed for the purpose of being used in the laboratory are highly reduced miniature languages. They are language-like in the sense that they carry basic features of natural languages: they consist of syllables that are composed of consonants and vowels, they may carry prosodic information, but they could never be used for communication. In spite of this extreme reduction, humans have been found to process ALs by relying on the same language learning and processing mechanisms that they rely on when learning or processing natural languages. AL segmentation experiments can be conducted with infants and adults under the prerequisite that the procedure is adapted to the population of interest. Experiments with older children are also possible (Evans, Saffran, and Robe-Torres 2009), but relatively sparse.

The procedure in AL segmentation experiments consists of two phases. Experiments begin with a familiarization phase that may last for a minute but even up to 20 minutes. During this phase participants are supposed to learn the AL. Consecutively, there is a test phase in which is tested whether participants succeeded at learning the AL or which aspects of the language were learned. In the test phase, usually, items are set in contrast that occurred as “words” or as “partwords” in the AL. For an illustration, an excerpt of AL might be a sequence such as ...*fonagipezaro*... This AL can be segmented into the disyllables ...*fona* # *gipe* # *zaro* #... If this is the predicted segmentation, the researcher would refer to *fona*, *gipe*, and *zaro* as “words”. Formally, a segmentation of the other disyllables is also possible: ...*fo* # *nagi* # *peza* # *ro*... These disyllables are referred to as “partwords” (e.g., Saffran, Newport, and Aslin 1996a, 1996b). Partwords are contained in the speech stream but they straddle a word boundary. When testing adults on their segmentation abilities, they usually hear a pair of a word and a partword, and their task is to decide which of the two was a word of the AL they just heard. Segmentation preferences are evaluated by analyzing the responses.

Infants can be tested as follows: after familiarization with the AL, they hear trials that either consist of words or of partwords. The head-turn preference procedure (Kemler Nelson et al. 1995) is used to measure segmentation preferences. In this procedure, test stimuli are randomly played from a loudspeaker installed at the left or the right side of the experimental booth. Both loudspeakers are coupled with a lamp installed next to the loudspeaker to enhance the infants’ attraction to turn the head towards the speaker. If infants prefer listening to one type of stimulus over the other (measured in looking times), this is interpreted to be indicative of their segmentation ability. Usually, however, it cannot be predicted whether infants prefer words that they have segmented from the language (familiarity effect) or partwords that occur as new to them (novelty effect). In order to interpret results of segmentation experiments with infants, the community of infant language researchers relies on predictions by a model (Hunter and Ames 1988) that predicts novelty preferences if the processing task is relatively easy (i.e., the preference for the novel is a reflection of excitement after boredom as a result of habituation), whereas familiarity preferences are predicted if the task is relatively difficult for infant participants. Factors that can influence whether novelty or familiarity effects are yielded are, for example, the infants’ age, the complexity of the linguistic stimuli, and the length of the familiarization phase.

3.2 Segmentation tasks with naturalistic stimuli

Adults' segmentation abilities can be tested in the word spotting task (first used by Cutler and Norris 1988) that is similar to a lexical decision task. In word spotting, participants hear nonwords, and these nonwords may or may not contain a real word. The participants' task is to listen carefully and press a button if they detect a real word in the nonsense speech. Since the words to be spotted are embedded in a speech context, it is generally accepted that the task tests speech segmentation. Words can be embedded at the beginning or at the end of a stimulus, and usually, the context in which the word occurs is manipulated to test the effect of potential cues for speech segmentation. To assess participants' segmentation abilities, reaction times are typically analyzed.

In its original sense, the word spotting task cannot be used for cross-linguistic comparisons, as, in the ideal case, experiments designed for comparing speakers of different languages require that the same stimuli are used across groups, and obviously, it is impossible to spot words that are not part of one's language. However, newer studies have found solutions for circumnavigating this problem. Lentz (2011) used a word-spotting task combined with eye-tracking of looks to letters on screen, and phonotactic boundary cues directed participants' looks to letters in a similar way independent of whether the "target" item was a word or a nonword. The fact that nonword items show the effect of phonotactics even stronger than words offers new possibilities to compare speech segmentation cross-linguistically. Moreover, recent cross-linguistic studies (Kabak, Maniwa, and Kazanina 2010; van Ommen 2016) have used a variant of the word spotting task in which nonwords had to be spotted in nonword contexts. Kabak, Maniwa, and Kazanina (2010) administered this by presenting the nonword targets orthographically on screen, and van Ommen (2016) first trained participants on the nonword targets. For a brief but detailed description of word spotting as a method, read McQueen (1996).

Infants' segmentation preferences can also be tested in a more naturalistic setting (e.g., Jusczyk and Aslin 1995; Jusczyk, Hohne, and Bauman 1999; Jusczyk, Houston, and Newsome 1999), again in the head-turn paradigm. Classically, in these studies, infants are familiarized with trials containing repetitions of isolated words. In the following test phase, infants hear text passages that either contain the words they were familiarized with or not. Looking times indicate if infants prefer one type of text passage over the other. In an adapted version of the classical procedure, infants are familiarized with text passages and tested with isolated words that were contained in the passages or not. Both variants are considered to be indicative of infants' segmentation ability, as they require that infants segment isolated words from speech, independent of whether the segmentation task itself has to be performed during familiarization or at test. In this task, segmentation performance can even be improved by means of an extended familiarization phase, in which parents regularly play text passages containing a target word to their infants at home before they are tested in the lab (Schreiner, Altvater-Mackensen, and Mani 2016). The task has successfully been applied with non-native material and used for cross-linguistic comparisons (Houston et al. 2000; Polka and Sundara 2012). Moreover, we have recently found that the task can also be used with more controlled nonsense speech text passages and non-word targets (Boll-Avetisyan, Fritzsche, and Jäkel 2017).

3.3 Comparing the methodological advantages of the different segmentation tasks

3.3.1 Advantages of word spotting

Advantages of the word spotting task over the AL segmentation task is that it provides a relatively natural speech segmentation condition. Another advantage is that it allows for directly assessing the influence of the context on word recognition, while experiments with ALs involve other factors like memory, as segmentation is only assessed after familiarization. Moreover, word spotting bears the advantage of having access to an online measure of segmentation performance by measuring reaction times. This online measure may be more sensitive to some effects on segmentation performance than response preferences, the offline measure analyzed in AL segmentation studies.

3.3.2 Advantages of tasks with naturalistic stimuli in infants

Infant experiments are generally difficult to test and it is a challenging task to find an experimental procedure that works well and reveals infants' language processing abilities. Hence, one advantage of segmentation tasks with naturally recorded stimuli is that this procedure has been used in numerous studies (compared to relatively fewer infant studies using ALs), and the community of infant language researchers appreciates that these experiments produce solid, replicable results.

3.3.3 Advantages of AL segmentation

Researchers like using ALs as experimental stimuli, as they have the compelling benefit of being highly controllable for confounds while testing the cue of interest in isolation (specific examples of how confounding factors can be controlled are given in Section 3.4). Usually, AL experiment materials are synthesized using text-to-speech software (e.g., MBROLA by Dutoit et al. 1996). By this, it is possible to generate long speech streams without pauses that contain co-articulatory cues between all adjacent segments. This degree of control for phonetic cues for speech segmentation could never be reached with a human speaker. Furthermore, it is possible to control for an influence of any sub-lexical cue by, for example, keeping segment durations, intensity and pitch constant across syllables. Moreover, ALs do not include real words, hence, an influence of lexical representations can be minimized. Furthermore, if designed appropriately, the same stimuli can be used to compare native speakers of different languages or to compare infants and adults.

For those who are interested in the question of which phonological knowledge is transferred when processing an L2 in the initial state of learning, AL segmentation experiments can also give insights. By minimizing many features that are characteristic of the L1, an AL is essentially a miniature version of an L2.

Another nice feature of AL segmentation experiments is that different segmentation cues can be put in conflict. For example, an AL can be constructed such that statistical distributions give cues to words, while prosody gives cues to partwords (as in, e.g., Johnson and Jusczyk 2001; Thiessen and Saffran 2003). An AL can also be constructed such that abstract phonotactic constraints cue to a different segmentation than phonotactic detail that is exempt from the general effect of the abstract phonotactic constraint (as in, e.g., Boll-Avetisyan and Kager 2016). This feature, which will be difficult to add into a word spotting experiment, is particularly valuable for researchers who are interested in exploring how listeners weight different cues at simultaneous occurrence in speech segmentation. If offers many interesting possibilities for future experiments, for example for contrasting L1 versus L2 cues, language-specific versus universal cues.

3.4 Methodological recommendations for artificial language segmentation experiments

Above, I have briefly sketched out some of the advantages of the AL segmentation task. In my view, the AL segmentation paradigm is ideal for addressing some of the questions that an experimental phonologist may ask. For example, if a researcher is interested in the influence of language-specific versus language-general (universal) phonological cues for speech segmentation, it will be relatively simple to design an AL that puts both cues in conflict. Furthermore, the AL could be designed such that the same AL material is used with infants or adults with different language backgrounds to get a better understanding of how language-specific versus universal properties of phonological words may interact in acquisition and processing. All this can be achieved by carefully controlling for confounds from other sub-lexical or lexical cues for speech segmentation. The controlling of confounding factors is extremely important for avoiding the risk of misinterpreting effects on speech segmentation. The following section provides some methodological recommendations for planning AL segmentation experiments.

3.4.1 Material

AL stimuli usually consist of an artificial speech stream, which are made up of words (and partwords). It needs to be considered which characteristics the words should carry. First, a syllabic structure needs to be selected, which, ideally, should be the same for all words and part words. Most studies have made up their words from simple CV syllables (if the researcher wants to assess how more complex syllable structures are segmented, this is also possible; see Ettlinger, Finn, and Hudson Kam 2012; Finn and Hudson Kam 2008). Second, the number of syllables contained in a word needs to be decided on. Monosyllabic words are not used because the classical design involves a comparison between "words" and "partwords" at test. Most studies use di- or trisyllabic words, as these patterns are frequent in natural languages, and relatively easy to memorize. Third, specific consonants and vowels (i.e., the syllables) need to be selected and brought into an ideal order. When cross-linguistic comparisons are planned, all phonemes should be part of the phoneme inventories of the participants' native languages. In order to control for an influence of the L1(s), it can be recommended to select syllables such that they are balanced in terms of sub-lexical cues to segmentation that are not of interest to the researcher. Otherwise researchers run into risk that their results are confounded (see Onnis et al. 2005, who show that such a confound could explain Peña et al.'s 2002 results). Sub-lexical cues may come from positional segment, biphone, and syllable probabilities, cohort size, or lexical neighborhood size, and it is possible to control for these factors (see Boll-Avetisyan 2012; Boll-Avetisyan and Kager 2014; Boll-Avetisyan and Kager

2016). Fourth, the researcher should control for prosodic cues to segmentation by flattening the pitch and intensity information on the whole stimulus, and by controlling for segment duration.

3.4.2 Procedure

With regards to the experimental procedure, the researcher has to consider the following. First, the length of the familiarization phase has to be determined. While earlier studies often used very long familiarization phases of up to 20 minutes, later studies have used much shorter versions (e.g., 3 minutes). From experience, if participants are exposed to long familiarization streams, they sometimes report that they first heard a number of specific words and all of a sudden had the impression that a whole bunch of new words entered the stream. From these reports it is evident that they first heard "words", but switched to hearing the "partwords" at some point. Hence, shorter familiarization phases may be preferable. Moreover, note that adults' performance in AL segmentation experiments is better if their attention during familiarization is not directed to an unrelated task (like coloring mandalas, Toro, Sinnett, and Soto-Faraco 2005).

Second, the presentation of items in the test phase has to be decided. Traditionally, studies use two-alternatives forced choice tasks, in which a trial consists of a word and a partword. It is also possible to present just one item at a time, and ask participants, whether this item was part of the AL or not (e.g., Bion, Benavides-Varela, and Nespor 2011). The disadvantage of this procedure is that fewer items of one type can be used.

One should avoid long test phases, as there is evidence that memory of the words learned from the AL decays relatively soon, once participants are flooded with a lot of totally "new" items (that they did not segment from the AL) during the test phase. This memory decay over the test phase has been observed in AL segmentation experiments with both adults (Boll-Avetisyan and Kager 2014) and infants (Abboub et al. 2016). Note that a decay of memory for novel items acquired from brief exposure is not unique to the AL segmentation paradigm. We also see this in other experiments that use a familiarization plus test phase procedure (e.g., Boll-Avetisyan, Fritzsche, and Jäkel 2017, where we familiarized infants with isolated words and tested their segmentation with text passages). As test phases with few trials may go on the expense of statistical power, the researcher may nonetheless opt for a longer test phase and plan to include "trial number" as a factor in the statistical analysis.

3.5 Synthesis

In the present section, I have introduced the different experimental tasks that can be used for testing effects of phonological knowledge on speech segmentation. Next, I summarized some of the main benefits of the different tasks. Attention was drawn to the fact that the AL segmentation task has particular benefits for the experimental phonologist with an interest in effects of the properties of phonological words on segmentation. These were the possibilities to a) control the material for confounding factors *ad extremum*, b) use stimuli that do not involve real words, c) create stimuli that will be processed in a similar way by groups of speakers of different languages, and d) put different cues in conflict. Hence, in the following, recommendations were offered for the design of stimulus material and experimental procedure in AL segmentation experiments.

4 Concluding remarks

In the first part of this chapter, it was demonstrated that restrictions on the phonological structure of words have consequences for the phonological structure of speech, and that this phonological structure provides valuable cues for speech segmentation for infants and adults. The second part of the chapter focused on experimental methods that can be applied when testing effects of phonological knowledge on infants' and adults' speech segmentation. Special attention was given to the AL segmentation task as a method, which, as it was argued, bears some particular advantages for addressing research questions of interest to experimental phonologists.

One of the questions that have always been in the center of interest of many phonologists is whether phonological structure reflects effects of innate biases. I suggest that when combining these benefits of the AL segmentation paradigm in one experimental study, the method is ideal for finding empirical support for effects of innate biases on processing. The first crucial step is to select a phenomenon of interest, that is, a phonological structure or a phonological constraint that is supposedly innate. Second, one would have to find a language, in which words are not affected by this phonological constraint. Infants and adults can then be tested in an AL segmentation experiment, in which the supposedly innate phonological constraint provides a cue to word boundaries. Potentially, this cue can be put in conflict with another cue that would not be universally preferred in order to have a direct comparison and potentially additional evidence for an effect of cues that are universally preferred. Ideally, the experiment would include a control group of speakers of a language in which the assumed phonological bias affects the lexicon. Such experimental studies could get us to the heart of gaining an understanding of which underlying phonological representations of the structure of phonological words influence processing.

References

Abboub, Nawal, Natalie Boll-Avetisyan, Anjali Bhatara, Barbara Höhle & Thierry Nazzi. 2016. An exploration of rhythmic grouping of speech sequences by French- and German-learning infants. Frontiers in Neurosciences 10. 292.

Adriaans, Frans & René Kager. 2010. Adding generalization to statistical learning: The induction of phonotactics from continuous speech. Journal of Memory and Language 62. 311–331.

Babineau, Mireille & Rushen Shi. 2014. Distributional cues and the onset bias in early word segmentation. Developmental Psychology 50. 2666–2674.

Bergmann, Pia. 2018. The phonological word in German – Insights from an acoustic-phonetic study of complex words. In Christiane Ulbrich, Alexander Werth & Richard Wiese (eds.), Empirical approaches to the phonological structure of words (Linguistische Arbeiten 567), 13–44. Berlin & Boston: De Gruyter.

Bhatara, Anjali, Natalie Boll-Avetisyan, Annika Unger, Thierry Nazzi & Barbara Höhle. 2013. Native language affects rhythmic grouping of speech. Journal of the Acoustical Society of America 134(5). 3828–3843.

Bion, Ricardo A. H., Silvia Benavides-Varela & Marina Nespor. 2011. Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. Language and Speech 54. 123–140.

Blevins, Juliette. 1995. The syllable in phonological theory. In John Goldsmith (ed.), Handbook of Phonological Theory, 206–244. Oxford: Blackwell.

Boll-Avetisyan, Natalie. 2012. Phonotactics and its Acquisition, Representation, and Use: An Experimental-Phonological Study (LOT Dissertation series 298). Utrecht: LOT Publications.

Boll-Avetisyan, Natalie, Anjali Bhatara, Annika Unger, Thierry Nazzi & Barbara Höhle. 2016. Effects of experience with L2 and music on rhythmic grouping by French listeners. Bilingualism: Language and Cognition 19. 971–986.

Boll-Avetisyan, Natalie, Tom Fritzsche & Carolin Jäkel. 2017. German-learning infants recognize vowel-initial words in continuous speech at 11 months. Paper presented at the 3rd Workshop on Infant Language Development. Bilbao, Spain.

Boll-Avetisyan, Natalie & René Kager. 2014. OCP-Place in speech segmentation. Language and Speech 57. 394–421.

Boll-Avetisyan, Natalie & René Kager. 2016. Is speech processing influenced by abstract or detailed phonotactic representations? The case of the Obligatory Contour Principle. Lingua 171. 74–91.

Boll-Avetisyan, Natalie, René Kager, Elise de Bree & Annemarie Kerkhoff. 2009. The Obligatory Contour Principle in speech segmentation in pre-lexical infants. Paper presented at the KNAW Colloquium on Language Acquisition and Optimality Theory.

Brent, Michael R. & Timothy A. Cartwright. 1996. Distributional regularity and phonotactic constraints are useful for segmentation. Cognition 61. 93–125.

Cairns, Paul, Richard Shillcock, Nick Chater & Joe Levy. 1997. Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. Cognitive Psychology 33. 111–153.
Cole, Ronald A. & Jola Jakimik. 1980. A model of speech perception. In R. A. Cole (ed.), Perception and production of fluent speech, 133–163. Hillsdale, NJ: Erlbaum.
Curtin, Suzanne, Toben Mintz & Morten H. Christiansen. 2005. Stress changes the representational landscape: evidence from word segmentation. Cognition 96. 233–262.
Cutler, Anne. 1994. Segmentation problems, rhythmic solutions. Lingua 92. 81–104.
Cutler, Anne & Sally Butterfield. 1992. Rhythmic cues to speech segmentation: Evidence from juncture misperception. Journal of Memory and Language 31. 218–236.
Cutler, Anne & Jacques Mehler. 1993. The periodicity bias. Journal of Phonetics 21. 103–108.
Cutler, Anne, Jacques Mehler, Dennis Norris & Juan Segui. 1989. Limits on bilingualism. Nature 340. 229–230.
Cutler, Anne, Jacques Mehler, Dennis Norris & Juan Segui. 1992. The monolingual nature of speech segmentation by bilinguals. Cognitive Psychology 24. 381–410.
Cutler, Anne & Dennis Norris. 1988. The role of strong syllables in segmentation for lexical access. Journal of Experimental Psychology: Human Perception and Performance 14. 113–121.
Domahs, Ulrike, Frank Domahs & Christina Kauschke. 2018. The morphology-prosody interface in typically developing and language-impaired populations: An overview and insights from experimental approaches. In Christiane Ulbrich, Alexander Werth & Richard Wiese (eds.), Empirical approaches to the phonological structure of words (Linguistische Arbeiten 567), 95–120. Berlin & Boston: De Gruyter.
Dutoit, Thierry, Vincent Pagel, Nicolas Pierret, Frédéric Bataille & Olivier van der Vreken. 1996. The MBROLA project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. Paper presented at Proceedings of The Fourth International Conference on Spoken Language Processing (ICSLP'96) 3.
Ettlinger, Marc, Amy S. Finn & Carla L. Hudson Kam. 2012. The effect of sonority on word segmentation: Evidence for the use of a phonological universal. Cognitive Science 36. 655–673.
Evans, Julia L., Jenny R. Saffran & Kathryn Robe-Torres. 2009. Statistical learning in children with specific language impairment. Journal of Speech, Language, and Hearing Research 52. 321–335.
Féry, Caroline, Robin Hörnig & Serge Pahaut. 2011. Correlates of phrasing in French and German from an experiment with semi-spontaneous speech. In Christoph Gabriel & Conxita Lleó (eds.), Intonational Phrasing in Romance and Germanic: Cross-linguistic and Bilingual Studies, 11–42. Amsterdam: Benjamins.
Finn, Amy S. & Carla L. Hudson Kam. 2008. The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. Cognition 108. 477–499.
Frost, Rebecca L. A., Padraic Monaghan & Tomoko Tatsumi. 2017. Domain-general mechanisms for speech segmentation: The role of duration information in language learning. Journal of Experimental Psychology: Human Perception and Performance 43(3). 466–476.
Gleitman, Lila R. & Eric Wanner. 1982. Language acquisition: The state of the state of the art. In Eric Wanner & Lila R. Gleitman (eds.), Language acquisition: The state of the art, 3–48. Cambridge, MA: Cambridge University Press.
Goldsmith, John. 1976. An overview of autosegmental phonology. Linguistic Analysis 2. 23–68.

Gonzalez-Gomez, Nayeli & Thierry Nazzi. 2013. Effects of Prior Phonotactic Knowledge on Infant Word Segmentation: The Case of Nonadjacent Dependencies. Journal of Speech, Language, and Hearing Research 56. 840–849.
Goyet, Louise, Scania de Schonen & Thierry Nazzi. 2010. Words and syllables in fluent speech segmentation by French-learning infants: An ERP study. Brain Research 1332. 75–89.
Hall, T. Alan. 1999. The phonological word: A review. In T. Alan Hall & Ursula Kleinhenz (eds.), Studies on the Phonological Word, 1–22. Amsterdam & Philadelphia: Benjamins.
Harris, Zellig S. 1944. Simultaneous components in phonology. Language 20. 181–205.
Hay, Jessica F. & Jenny R. Saffran. 2012. Rhythmical grouping biases constrain infant statistical learning. Infancy 17. 610–641.
Hayes, Bruce. 1995. Metrical stress theory: principles and case studies. Chicago & London: The University of Chicago Press.
Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. Linguistic Inquiry 39. 379–440.
Höhle, Barbara, Ranka Bijeljac-Babic, Birgit Herold, Jürgen Weissenborn & Thierry Nazzi. 2009. Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. Infant Behavior and Development 32. 262–274.
Houston, Derek M., Peter W. Jusczyk, Cecile Kuijpers, Riet Coolen & Anne Cutler. 2000. Cross-language word segmentation by 9-month-olds. Psychonomic Bulletin & Review 7. 504–509.
Hunter, Michael A. & Elinor W. Ames. 1988. A multifactor model of infant preferences for novel and familiar stimuli. In Carolyn Rovee-Collier & Lewis P. Lipsitt (eds.), Advances in infancy research, 69–95. Westport, CT: Ablex Publishing.
Johnson, Elizabeth K. & Peter W. Jusczyk. 2001. Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics. Journal of Memory and Language 44. 548–567.
Johnson, Elizabeth K., Amanda Seidl & Michael D. Tyler. 2014. The edge factor in early word segmentation: Utterance-level prosody enables word form extraction by 6-month-olds. PLoS ONE 9. e83546.
Jusczyk, Peter W. 1997. The discovery of spoken language. Cambridge, MA: MIT Press.
Jusczyk, Peter W. & Richard N. Aslin. 1995. Infants' detection of sound patterns of words in fluent speech. Cognitive Psychology 29. 1–23.
Jusczyk, Peter W., Elizabeth Hohne & Angela Bauman. 1999. Infants' sensitivity to allophonic cues for word segmentation. Perception and Psychophysics 61. 1465–1467.
Jusczyk, Peter W., Derek M. Houston & Mary Newsome. 1999. The beginnings of word-segmentation in English-learning infants. Cognitive Psychology 39. 159–207.
Kabak, Baris, Kazumi Maniwa & Nina Kazanina. 2010. Listeners use vowel harmony and word-final stress to spot nonsense words: A study of Turkish and French. Journal of Laboratory Phonology 1. 207–224.
Kemler Nelson, Deborah G., Peter W. Jusczyk, Denise R. Mandel, James Myers, Alice Turk & LouAnn Gerken. 1995. The Head-Turn Preference Procedure for testing auditory perception. Infant Behavior and Development 18. 111–116.
Kooijman, Valesca, Peter Hagoort & Anne Cutler. 2009. Prosodic Structure in Early Word Segmentation: ERP Evidence From Dutch Ten-Month-Olds. Infancy 14. 591–612.
Lasky, Robert E., Ann Syrdal-Lasky & Robert E. Klein. 1975. VOT discrimination by four to six and a half month old infants from Spanish environments. Journal of Experimental Child Psychology 20. 215–225.

Lentz, Tom O. 2011. Phonotactic Illegality and Probability in Speech Perception: Evidence from Second Language Listeners (LOT Dissertation Series 290). Utrecht: LOT.
Liberman, Mark & Alan Prince. 1977. On stress and linguistic rhythm. Linguistic Inquiry 8. 249–336.
Marslen-Wilson, William D. & Alan Welsh. 1978. Processing interactions and lexical access during word recognition in continuous speech. Cognitive Psychology 10. 29–63.
Mattys, Sven L. & Peter W. Jusczyk. 2001a. Phonotactic cues for segmentation of fluent speech by infants. Cognition 78. 91–121.
Mattys, Sven L. & Peter W. Jusczyk. 2001b. Do infants segment words or recurring contiguous patterns? Journal of Experimental Psychology: Human Perception and Performance 27. 644–655.
Mattys, Sven L., Peter W. Jusczyk, Paul A. Luce & James L. Morgan. 1999. Phonotactic and prosodic effects on word segmentation in infants. Cognitive Psychology 38. 465–494.
Mattys, Sven L., Laurence White & James F. Melhorn. 2005. Integration of Multiple Speech Segmentation Cues: A Hierarchical Framework. Journal of Experimental Psychology: General 134. 477–500.
McQueen, James. 1996. Word Spotting. Language and Cognitive Processes 11. 695–699.
McQueen, James. 1998. Segmentation of Continuous Speech Using Phonotactics. Journal of Memory and Language 39. 21–46.
Mehler, Jacques, Jean Yves Dommergues, Uli Frauenfelder & Juan Segui. 1981. The syllable's role in speech segmentation. Journal of Verbal Learning and Verbal Behavior 20. 298–305.
Mehler, Jacques, Peter W. Jusczyk, Ghislaine Lambertz, Nilofar Halsted, Josiane Bertoncini & Claudine Amiel-Tison. 1988. A precursor of language acquisition in young infants. Cognition 29. 143–178.
Moreton, Elliott. 2008. Analytic bias and phonological typology. Phonology 25. 83–127.
Moreton, Elliott & Joe Pater. 2012. Structure and substance in artificial-phonology learning. Part I, Structure. Part II, Substance. Language and Linguistics Compass 6. 686–701 and 702–718.
Morgan, James L. & Katherine Demuth. 1996. Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition. Mahwah, NJ: Erlbaum.
Nazzi, Thierry, Josiane Bertoncini & Jacques Mehler. 1998. Language discrimination by newborns: Toward an understanding of the role of rhythm. Journal of Experimental Psychology: Human Perception and Performance 2. 756–766.
Nazzi, Thierry, Laura C. Dilley, Ann Marie Jusczyk, Stefanie Shattuck-Hufnagel & Peter W. Jusczyk. 2005. English-learning infants' segmentation of verbs from fluent speech. Language and Speech 48. 279–298.
Nazzi, Thierry, Galina Iakimova, Josiane Bertoncini, Séverine Frédonie & Carmela Alcantara. 2006. Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. Journal of Memory and Language 54. 283–299.
Nazzi, Thierry, Karima Mersad, Megha Sundara, Galina Iakimova & Linda Polka. 2014. Early word segmentation in infants acquiring Parisian French: task-dependent and dialect-specific aspects. Journal of Child Language 41. 600–633.
Nespor, Marina & Irene Vogel. 1986. Prosodic phonology. Dordrecht: Foris.
Nespor, Marina, Mohinish Shukla, Ruben van de Vijver, Cinzia Avesani, Hanna Schraudolf & Caterina Donati. 2008. Different phrasal prominence realizations in VO and OV languages. Lingue e linguaggio 7. 1–28.

Norris, Dennis, James McQueen & Anne Cutler. 1995. Competition and segmentation in spoken-word recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition 21. 1209–1228.

Ommen, Sandrien van. 2016. Listen to the beat: A cross-linguistic perspective on the use of stress in segmentation (LOT Dissertation Series 420). Utrecht: LOT.

Onnis, Luca, Padraic Monaghan, Korin Richmond & Nick Chater. 2005. Phonology impacts segmentation in speech processing. Journal of Memory and Language 53. 225–237.

Pelucchi, Bruna, Jessica F. Hay & Jenny R. Saffran. 2009. Learning in reverse: 8-month-old infants track backward transitional probabilities. Cognition 113. 244–247.

Peña, Marcela, Luca Bonatti, Marina Nespor & Jacques Mehler. 2002. Signal-driven computations in speech processing. Science 298. 604–607.

Peperkamp, Sharon. 1999. Prosodic words. Paper presented at GLOT International 4.4.

Perruchet, Pierre & Annie Vinter. 1998. PARSER: A model for word segmentation. Journal of Memory and Language 39. 246–263.

Peters, Ann M. 1985. Language segmentation: Operating principles for the perception and analysis of language. In Dan I. Slobin (ed.), The crosslinguistic study of language acquisition, 1029–1068. Hillsdale, New Jersey: L. Erlbaum Associates.

Pierrehumbert, Janet B. 2003. Phonetic diversity, statistical learning, and acquisition of phonology. Language and Speech 46. 115–154.

Polka, Linda & Megha Sundara. 2012. Word segmentation in monolingual infants acquiring Canadian English and Canadian French: Native language, cross-dialect, and cross-language comparisons. Infancy 17. 198–232.

Ramus, Franck, Marina Nespor & Jacques Mehler. 1999. Correlates of linguistic rhythm in the speech signal. Cognition 73. 265–292.

Saffran, Jenny R., Elissa L. Newport & Richard N. Aslin. 1996a. Statistical learning by 8-month-old infants. Science 274. 1926–1928.

Saffran, Jenny R., Elissa L. Newport & Richard N. Aslin. 1996b. Word Segmentation: The Role of Distributional Cues. Journal of Memory and Language 35. 606–621.

Sandoval, Michelle & Rebecca L. Gómez. 2016. Overriding the metrical bias with lexical information: English-learning 7.5-month-olds use mommy to segment iambic words. Language Learning and Development 12. 398–412.

Scharinger, Mathias. 2018. Neural bases of phonological representations: Empirical approaches and methods. In Christiane Ulbrich, Alexander Werth & Richard Wiese (eds.), Empirical approaches to the phonological structure of words (Linguistische Arbeiten 567), 241–272. Berlin & Boston: De Gruyter.

Schreiner, Melanie S., Nicole Altvater-Mackensen & Nivedita Mani. 2016. Early word segmentation in naturalistic environments: Limited effects of speech register. Infancy 21. 625–647.

Seidl, Amanda & Elizabeth K. Johnson. 2008. Boundary alignment enables 11-month-olds to segment vowel initial words from speech. Journal of Child Language 35. 1–24.

Selkirk, Elisabeth O. 1982. The Syllable. In Harry van der Hulst & Norval Smith (eds.), The Structure of Phonological Representations, Part II, 337–383. Dordrecht: Foris.

Skoruppa, Katrin, Andrew Nevins, Adam Gillard & Stuart Rosen. 2015. The role of vowel phonotactics in native speech segmentation. Journal of Phonetics 49. 67–76.

Suomi, Kari, James McQueen & Anne Cutler. 1997. Vowel harmony and speech segmentation in Finnish. Journal of Memory and Language 36. 422–444.

Swingley, Daniel. 2005. Statistical clustering and the contents of the infant vocabulary. Cognitive Psychology 50. 86–132.

Tagliapietra, Lara, Rachele Fanari, Chiara De Candia & Patrizia Tabossi. 2009. Phonotactic regularities in the segmentation of spoken Italian. The Quarterly Journal of Experimental Psychology 62. 392–415.
Thiessen, Erik D. & Jenny R. Saffran. 2003. When Cues Collide: Use of Stress and Statistical Cues to Word Boundaries by 7- to 9-Month-Old Infants. Developmental Psychology 39. 706–716.
Toro, Juan M., Scott Sinnett & Salvador Soto-Faraco. 2005. Speech segmentation by statistical learning depends on attention. Cognition 97. B25–B34.
Tremblay, Annie, Mirjam Broersma, Caitlin E. Coughlin & Jiyoun Choi. 2016. Effects of the native language on the learning of fundamental frequency in second-language speech segmentation. Frontiers in Psychology 7. 985.
Trubetzkoy, Nikolai S. 1939/1969. Principles of phonology translated by C.A. Baltaxe. Berkeley, CA: University of California.
Tyler, Michael D. & Anne Cutler. 2009. Cross-language differences in cue use for speech segmentation. Journal of the Acoustical Society of America 126. 367–376.
Ulbrich, Christiane, Phillip M. Alday, Johannes Knaus, Paula Orzechowska & Richard Wiese. 2016. The role of phonotactic principles in language processing. Language, Cognition and Neuroscience 31. 662–682.
Ulbrich, Christiane & Richard Wiese. 2018. Phonotactic principles and exposure in second language processing. In Christiane Ulbrich, Alexander Werth & Richard Wiese (eds.), Empirical approaches to the phonological structure of words (Linguistische Arbeiten 567), 153–182. Berlin & Boston: De Gruyter.
Van Kampen, Anja, Güliz Parmaksiz, Ruben van de Vijver & Barbara Höhle. 2008. Metrical and statistical cues for word segmentation: Vowel harmony and word stress as cues to word boundaries by 6- and 9-month old Turkish learners. In Anna Gavarró & M. João Freitas (eds.), Language Acquisition and Development, 313–324. Newcastle: Cambridge Scholars Publishing.
Vroomen, Jean, Jyrki Tuomainen & Beatrice de Gelder. 1998. The roles of word stress and vowel harmony in speech segmentation. Journal of Memory and Language 38. 133–149.
Vroomen, Jean, Monique van Zon & Beatrice de Gelder. 1996. Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. Memory & Cognition 24. 744–755.
Warner, Natasha, Jeesun Kim, Chris Davis & Anne Cutler. 2005. Use of complex phonological patterns in speech processing: Evidence from Korean. Journal of Linguistics 41. 353–387.
Weber, Andrea & Anne Cutler. 2006. First-language phonotactics in second-language listening. Journal of the Acoustical Society of America 119. 597–607.
Weissenborn, Jürgen & Barbara Höhle (eds.). 2001. Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition. Philadelphia, PA: Benjamins.
White, James & Megha Sundara. 2014. Biased generalization of newly learned phonological alternations by 12-month-old infants. Cognition 133. 85–90.
Wiese, Richard, Paula Orzechowska, Phillip M. Alday & Christiane Ulbrich. 2017. Structural principles or frequency of use? An ERP experiment on the learnability of consonant clusters. Frontiers in Psychology 7. 2005.
Zahner, Katharina, Muna Schönhuber & Bettina Braun. 2016. The limits of metrical segmentation: intonation modulates infants' extraction of embedded trochees. Journal of Child Language 43. 1338–1364.

Mathias Scharinger

Neural bases of phonological representations: Empirical approaches and methods

Abstract: This chapter provides an overview of brain imaging methods for studying phonological representations. The focus is on studies that consider linguistically motivated assumptions about representations, assuming that the dominant perceptual unit is the phonological feature.

Keywords: phonological representations, brain imaging, superior temporal sulcus, underspecification

1 Introduction

Empirical approaches to phonology with a focus on the psychological and neurobiological bases of language comprehension commonly refer to a core perceptual unit that is labeled by principal terms of theoretical phonology, such as feature, phoneme, syllable, or word (Sendlmeier 1995), or by terms of cognitive psychology, such as exemplar or prototype (Rosch 1973). These approaches assume phonological representations with differing degrees of abstraction from the physical signal and the corresponding articulatory configurations. The bandwidth of abstraction reaches from very detailed, episodic representations (exemplars, e.g. Pierrehumbert 2002) to very sparse, minimal representations, described by means of phonological features (Jakobson 1939). Phonological features as sub-phonemic, contrastive units refer to specific acoustic properties concomitant to particular articulator configurations that can be expressed in a binary manner with e.g. ±voice distinguishing between voiced [d] and voiceless [t] (Chomsky and Halle 1968). Alternatively, features can be expressed in a privative manner with voiced sounds marked with [voice] and voiceless sounds entirely lacking this feature (Lombardi 1996). Phonological features are one way to describe invariant, higher-order acoustic cues or cue-relations in speech perception (Lahiri, Gewirth, and Blumstein 1984).

Mathias Scharinger: [1]Philipps-Universität Marburg, Pilgrimstein 16, D-35032 Marburg, mathias.scharinger@staff.uni-marburg.de; [2]Max Planck Institute for Empirical Aesthetics, Grüneburgweg 14, D-60322 Frankfurt am Main, mathias.scharinger@gmail.com

DOI 10.1515/9783110542899-010

The description of phonemic contrasts by means phonological features has a long tradition in linguistic theory (Fant 1960; Jakobson 1939) and recently received increased attention from neurobiological approaches to phonology (Cornell, Lahiri, and Eulitz 2011; Eulitz and Lahiri 2004; Phillips et al. 2000). This may have to do with two reasons:

First, the phonological feature as abstract and categorical perceptual unit seems to adequately describe aspects of the cortical processing hierarchy. Within this hierarchy, detailed acoustic (within-category) information is processed in primary auditory areas (Heschl's gyrus). On the other hand, more abstract, invariant (across-category) speech information as well as supposedly universal principles (e.g. phonotactic constraints derived from the sonority hierarchy) are processed in surrounding areas, most prominently superior temporal sulcus (Davis and Johnsrude 2003; Hickok and Poeppel 2007; Obrig, Mentzel, and Rossi 2016; Okada et al. 2010; Poeppel, Idsardi, and van Wassenhove 2008; Scott and Johnsrude 2003). Note that this division of labor between these specific brain areas pays tribute to the observation that phonological processing may involve aspects of both abstract principles and detailed input patterns (see Ulbrich and Wiese 2018, this volume).

Second, phonological features provide an elegant way of describing phonological asymmetries and their perceptual consequences by referring to different degrees of representational specificity (underspecification, Lahiri and Reetz 2002, 2010).

Phonological underspecification, albeit not undisputed in phonological theory and psycholinguistic studies (McCarthy and Taub 1992), is based on the observation that some phonemes (e.g. those articulated with the tip or blade of the tongue, i.e. coronal sounds) are much more likely to undergo phonological processes such as assimilations (lea/n/ /b/acon > lea[m] bacon) while others (e.g. those articulated with the dorsum of the tongue, i.e. dorsal sounds) are not likely to undergo assimilations (plu/m/ /t/offee > *plu[n] toffee). As a consequence, coronal speech sounds are conceived of being underspecified for their place of articulation feature (no feature), while dorsal speech sounds are assumed to have a place of articulation specification (dorsal feature). In assimilation, then, underspecified coronal speech sounds can receive place of articulation features from neighboring sounds, while specified dorsal sounds cannot. In a recent psycho- and neurolinguistically defined version of underspecification (Lahiri and Reetz 2002, 2010), it is assumed that underspecified speech sounds can also be activated by acoustic signals that do entirely match their featural structure (no mismatch), such that assimilated coronals can still access their underlying forms (i.e. lea[m] bacon is still understood as lea/n/ bacon). By contrast, specified dorsal sounds cannot be activated by non-matching (mismatching) acoustic signals, such that plu[n] toffee cannot be interpreted as an assimilated version of plu/m/ toffee.

Underspecification as sparse coding principle has a strong affinity to coding principles in the neurosciences (Carlson, Ming, and DeWeese 2012). A particular resemblance seems to exist between underspecification and the concept of predictive coding (Baldeweg 2006; Friston 2005). Predictive coding refers to a framework of brain function wherein perception is not only based on sensory evidence but crucially on the brain's ability to generalize and to compute predictions against which bottom-up sensory evidence is compared. Predictions are derived from statistical priors (i.e. long-term memory information) and thus describe inferences about the most likely sensory environment. That is, in abstract terms, only non-redundant information is actively processed and represented in so-called error units (Friston 2010). These units represent the so-called prediction error that originates from mismatches between bottom-up sensory evidence and top-down predictions. Predictive coding has recently been discussed as representational framework (Gladziejewski 2015), and parallels between predictive coding and underspecification have been made in Scharinger et al. (2012) and Scharinger, Monahan, and Idsardi (2012), accounting for asymmetric effects in the electrophysiological investigation of speech sound perception.

Regarding the speech-is-special discussion (Carbonell and Lotto 2014), there seems to be a renewed interest in speech-specific cortical processing areas that are implicitly or explicitly assumed to house phonological representations (DeWitt and Rauschecker 2016; Hickok and Poeppel 2007; Obleser et al. 2007; Okada and Hickok 2006; Okada et al. 2010; Overath et al. 2015; Peelle, Johnsrude, and Davis 2010; Poeppel, Idsardi, and van Wassenhove 2008; Rauschecker and Scott 2009). Compared to previous, psycholinguistic investigations of phonological processing (e.g. Lahiri and Marslen-Wilson 1991; Scharinger and Lahiri 2010), the current methodological advances in the cognitive neurosciences allow for unprecedented accuracy in describing and analyzing the specificity of cortical regions for aspects of phonological representations and the time course of phonological processes during language comprehension.

This review tries to highlight and summarize the most important findings from the auditory neurosciences for illustrating and identifying the neural bases of phonological representations. Necessarily, this review cannot be exhaustive and will be restricted to (a) the auditory modality, (b) to phonological representations favoring the phoneme or the phonological feature and (c) to electrophysiological and metabolic measures of brain activity (EEG, MEG, fMRI) in healthy adults. In describing the most important methods and imaging measures applied in the field, this review may also serve as methodological guide for following up on specific research questions from phonological theory.

2 Overview of methods

Non-invasive brain imaging methods are used in order to better understand the functional relevance of cytoarchitectonically and anatomically defined brain regions, as well as their spatial and temporal characteristics underlying cognitive processes. Broadly speaking, brain imaging methods can be divided into measures with high spatial accuracy, most prominently exemplified by the localization of the hemodynamic response in functional magnet resonance imaging (fMRI) techniques, and into measures with high temporal accuracy, most prominently exemplified by the timing of electrophysiological measures such as electroencephalography (EEG) or magnetoencephalography (MEG).

2.1 Functional magnetic resonance imaging (fMRI)

The tool of fMRI for measuring the metabolism-related hemodynamic response is relatively young and was first used in the early 90s (Kwong et al. 1992; for details on the physics of the measure see Buxton 2009). In short, the measure of this technique is the blood oxygenation level dependent (BOLD) activity, conceived of as a correlate of local synaptic activity (Lee et al. 2010). Its underlying hemodynamic response is mainly brought about by the displacement of deoxyhemoglobin caused by inflowing oxygenated hemoglobin that is necessary for the neurons' energy supply. For this reason, the BOLD signal crucially relies on blood volume, blood flow and oxygen consumption in the brain. All these factors are most often positively correlated with neural activity (Logothetis and Wandell 2004).

Usually, fMRI studies rely on differences in BOLD activations between two conditions of interest. One of these conditions is either a resting-state (or silence) baseline during which the brain is in an idling mode, or a higher-level baseline that allows for a functional characterization of those areas that show more activation than the baseline. In fMRI experiments on phonological processing, a preferred baseline consists of a control task with non-speech (e.g. tone) stimuli. For instance, some experiments compare activations from phoneme discrimination with activations from tone discrimination (e.g. Humphries et al. 2014). Areas that show stronger BOLD signals in the phoneme than in the tone discrimination task can then be interpreted as being functionally specific to phoneme discrimination, rather than to auditory discrimination in general. For a complete picture, it is informative to also compare individual task activations to a resting-state or silence baseline. Thereby, it would be possible to identify not only areas that yield more activation for the phonological than for the auditory condition, but also overlapping phonological and auditory areas, i.e. areas where there is more activation in the phonological and auditory condition, compared to the resting state or silence condition. However, a comparison between a condition-of-interest (with task) and a resting-state condition (with no task) is confounded by task-specific activations when interpreting differences in activation. These differences may then just indicate that there is a task in one condition while there is no task in the other condition. In order to circumvent this confound, it is common to introduce a certain percentage of silent trials for which participants ideally provide responses similar to trials of interest (Liebenthal et al. 2013, for a discussion, see Caplan 2009).

Spatial resolution in fMRI experiments depends on the magnetic field strength of the scanner. For experiments involving humans, typical field strengths vary between 1.5 and 3 Tesla (T). Scanners with 3 T allow for a one-dimensional resolution of about 1 mm (or a voxel of 1 mm^3).

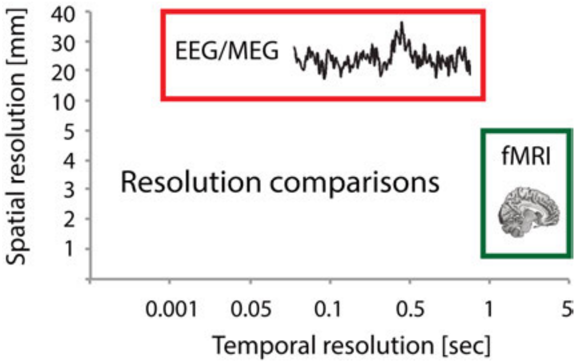


Fig. 1: Comparison of typical resolutions achieved by electrophysiological (EEG/MEG) and hemodynamic (fMRI) brain imaging methods. Electrophysiological methods offer high temporal resolution (on the order of 1 ms), while hemodynamic methods offer high spatial resolution (on the order of 1 mm). Scaling on x- and y-axis serves illustration only.

2.2 Electroencephalography (EEG) and Magnetoencephalography (MEG)

The methods of EEG and MEG can be summarized under electrophysiological methods because the key measure here is postsynaptic electric activity by which neurons operate and exchange information. EEG and MEG thus measure the ongoing electric activity of the brain. The EEG directly picks up the electric activity of several thousands of neurons by placing electrodes on the surface of the scalp, while MEG measures the concurrent changes in magnetic activity (for details, see Hansen, Kringelbach, and Salmelin 2010). The methods of EEG and MEG are largely complementary: EEG is better in picking up activity from radially oriented sources, while MEG is more sensitive to tangential sources. EEG is certainly a cheaper and more versatile method with the possibility of application in fieldwork. MEG, on the other hand, depends on a stationary scanner, but allows for a better source-localization because magnetic signal strength only decreases with the square of the distance from which it is measured, while the electric signal is affected by differing conductivities of bones and brain tissue in an unpredictable way.

Electrophysiological methods commonly focus on event-related potentials (ERPs in EEG) or event-related fields (ERFs in MEG). These notions refer to the averaging of time-locked responses to a large number of stimulus presentations. The rationale behind ERPs/ERFs (collectively called ERPs for the remainder of this chapter) is that an average of many time-locked responses to a specific event greatly improves the signal-to-noise ratio of the event-related response, averaging out any type of brain activity of no interest ("noise", see Molfese, Molfese, and Kelly 2001 for more details). Responses to language stimuli yield specific positive and negative deflections at different time points of the ERP, measured from stimulus onset. These deflections, differentially labeled positivities and negativities, have received linguistically-informed functional interpretations (Kutas, van Petten, and Kluender 2006).

3 Functional imaging experiments on phonological representations and processing

The majority of functional imaging experiments that explicitly focus on phonological aspects of speech perception try to demarcate brain regions that respond to aspects of stimulus properties beyond their acoustic shape, i.e. regions that are sensitive to abstraction from acoustic variation (cf. Obleser and Eisner 2009). It has been noted that this abstraction from acoustic variation has a spatial parallel in the human temporal lobes, in that more abstract (phonological) information is processed in regions with increasing distance from primary auditory areas (i.e. outside primary auditory cortex, an area within the lateral fissure and comprising parts of Heschl's gyrus, superior temporal gyrus, and planumtemporale; cf. Humphries et al. 2014; Obleser and Eisner 2009).

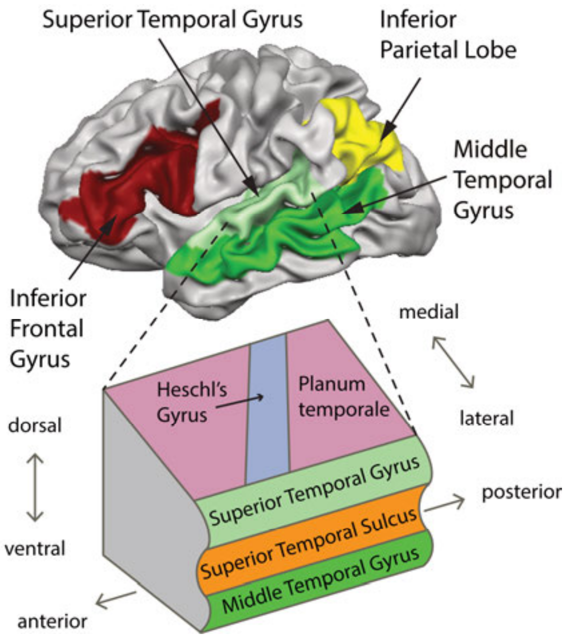


Fig. 2: Illustration of brain areas implicated in phonological processing. A schematic view shows details of auditory and phonological processing areas in the temporal lobes.

A key region for processing (and probably, housing) phonological information is the superior temporal sulcus, located between the superior and middle temporal gyrus in the human temporal lobes, and attributed a core function within the neural language network (Binder et al. 2000; Davis and Johnsrude 2007; Friederici 2012; Hickok and Poeppel 2007; Peelle, Johnsrude, and Davis 2010; Poeppel, Idsardi, and van Wassenhove 2008; Rauschecker and Scott 2009). While several studies have shown that the superior temporal sulcus supports the processing of intelligible speech (Davis and Johnsrude 2007; McGettigan et al. 2012; Obleser, Eisner, and Kotz 2008; Obleser and Kotz 2010; Scott et al.

2009; Scott et al. 2006), and the processing of speech sounds or phonetic aspects thereof (Binder et al. 2000; Turkeltaub and Coslett 2010), only a subset of imaging studies on speech and language explicitly mention its phonological function (illustrated in more detail in Table 1).

The majority of the reviewed studies focus on phonemic versus non-phonemic processing. For this reason, the stimulus material contains phonemic speech sounds or intelligible speech, and either acoustically-matched non-phonemic sounds or modified (e.g. spectrally rotated) speech material. In some cases, phonemic stimuli are contrasted with tone stimuli (with similar properties, e.g. duration or location of spectral peaks) in order to tease apart speech-specific abstract processing from acoustic processing. If tone stimuli are used, they are commonly labeled as control trials or baseline trials (see Table 1). Another aspect on which studies differ is whether they employ a passive listening task (with infrequent catch trials to maintain attention) or an active identification or discrimination task. Active tasks are most commonly same-different tasks of which the focus might be unspecific, or explicitly referring to acoustic or categorical stimulus properties (cf. Husain et al. 2006).

A very important aspect of auditory imaging studies concerns the scanner noise during continuous volume acquisitions that may mask the experimental stimuli. For this reason, some studies employ designs in which experimental stimuli are presented in-between two subsequent volume acquisitions, i.e. during silence. These designs are termed sparse (sometimes more detailed specified as ultrafast, interleaved, or clustered volume acquisition, cf. Peelle 2014; Schwarzbauer et al. 2006).

Determination of brain areas implicated in phonological processing may be achieved by standard analysis methods that contrast activity in a phonological condition to activity in a control or baseline condition (e.g. tone condition) for which the same task is required. Alternatively, different (phonological) conditions are compared to each other. The latter method is used in so-called repetition suppression paradigms, capitalizing on the fact that neurons adapt to stimulus repetitions or certain aspects of repeated stimulation. To this end, Vaden, Muftuler, and Hickok (2010) used three different types of phonological repetitions, high (identity; hip, hip), medium (cab, cat), and low (jug, knit) and then compared brain activations between low, medium and high repetition conditions. The assumption is that during perception, phonological areas would show graded repetition suppression, with strongest suppression in the high repetition condition and least suppression in the low repetition condition. The authors found this pattern to hold in bilateral superior temporal sulci.

Tab. 1: Overview of fMRI brain imaging studies on phonological processing during speech perception (auditory modality). Abbreviations: a–anterior, m–middle, p–posterior, STG–superior temporal gyrus, STS–superior temporal sulcus, MTG–middle temporal gyrus, PMC–pre-motor cortex, IFG–inferior frontal gyrus, SMG–supramarginal gyrus, IPL–inferior parietal lobe, PT–planumtemporale. Scanning type and methods are explained in the text.

Article	Task	Stimuli	Control	Scanning	Methods	Key Areas
Arsenault & Buchsbaum (2015)	speaker gender decision	English phonemes	silent trials	continuous	MVPA	aSTG
Ashtari et al. (2004)	same-different task	English words	tone trials, rest	continuous	standard	left MTG, STG
Buchsbaum, Hickok & Humphries (2001)	rehearsal	English pseudowords	tone trials, rest	continuous	standard	pSTG
Burton et al. (2005)	same-different task; rhyming detection	English pseudowords; words	tone trials; tone sequences	continuous	standard	left IFG, STG
Burton, Small & Blumstein (2000)	same-different task	English words [rhyming; non-rhyming]	tone trials	continuous	standard	left STG, MFG
Callan et al. (2004)	phoneme recognition	English syllables	silent trials	continuous	standard	aSTG, aSTS, MTG; IFG
Chevillet et al. (2013)	dichotic listening task	Synthetic da/ga continuum	silent trials	sparse	rapid adaptation	STG, PMC
Dehaene-Lambertz et al. (2005)	same-different task	Sinewave analog of ba/da continuum	silent trials	sparse	standard	left pSTS, SMG
Desai et al. (2008)	discrimination [ABX task]	Synthetic ba/da continuum; sinewave analog	silent trials	sparse	standard	left pSTS, STG
Deschamps, Baum & Gracco (2015)	passive listening; reading	English pseudowords, [differing sonority]	rest	sparse	standard	mSTS

Article	Task	Stimuli	Control	Scanning	Methods	Key Areas
Gelfand & Bookheimer (2003)	sequencing task	English syllables; hummed notes	implicit	continuous	standard	left IFG, SMG
Harinen & Rinne (2013)	discrimination; memory task	Finnish synthetic vowels, phonemic, nonphonemic	visual trials	continuous	standard	pSTG; IPL
Humphries et al. (2014)	passive listening; catch trials [silent]	English synthetic syllables, phonemic, nonphonemic	implicit	sparse	standard	STG, pSTS
Husain et al. (2006)	same-different task [acoustic, categorical]	English vowels, syllables; tones, tonal contours	rest	continuous	standard	IFG, IPL; STG
Jacquemot et al. (2003)	same-different task	Pseudowords; legal in French or Japanese	silent trials	sparse	standard	left pSTG, STS; aSMG
Lawyer & Corina (2014)	passive listening; catch trials [noise]	English syllables	catch trials	continuous	standard	STG, STS; MTG
Liebenthal et al. (2013)	identification [2AFC]	English syllables; chirps [duplex mode]	silent trials	sparse	standard	pSTG; IPL
Liebenthal et al. (2005)	discrimination [ABX task]	Synthetic ba/da continuum; nonphonemic analog	silent trials	sparse	standard	mSTS, aSTS; dorsal STG
Liebenthal et al. (2010)	identification [2AFC]	Synthetic ba/da continuum; nonphonemic analog	silent trials	sparse	standard	left mSTS; pSTS
LoCasto et al. (2004)	same-different task	English words/pseudowords; tone sequences; tones	implicit	continuous	standard	left IFG, MFG; left STG

Article	Task	Stimuli	Control	Scanning	Methods	Key Areas
McGettigan et al. (2010)	passive listening	English pseudowords [differing number of clusters; syllables]	tone trials, silent trials	sparse	standard	STG, PT
Obleser et al. (2006)	violation detection	German vowels [back, front, high, mid]	noise trials	sparse	standard	aSTG, STS
Okada et al. (2010)	intelligibility rating	English sentences, differing intelligibility	implicit	continuous	standard; MVPA	aSTS, pSTS
Okada & Hickok (2006)	passive listening	English pseudowords; words [differing density]	rest	continuous	standard	pSTS, mSTS
Oron et al. (2016)	rhyme detection (visual), initial consonant matching (auditory)	English words	tone trials	continuous	standard	mSTS (supramodal)
Rimol et al. (2005)	passive listening; catch trials [repetition]	Norwegian consonants, syllables; matched noises	noise trials	sparse	standard	MTG, pSTS
Scharinger et al. (2016)	same-different task	German pseudowords	silent trials/implicit	continuous	standard	aSTS, mSTS
Seghier et al. (2004)	rhyming detection [go/no-go]	French nouns [rhyming; non-rhyming]	implicit	continuous	standard	left pSTG, dorsal IFG
Strand et al. (2008)	same-different task	English pseudowords, [differing number of syllables]	noise trials [reversed speech]	continuous	standard	pSTS
Vaden, Muftuler & Hickok (2010)	passive listening; catch trials [pseudowords]	English words and pseudowords [differing repetitions]	implicit	continuous	standard	mSTS, pSTS

A method that is suggested to provide a better accuracy in distinguishing functional brain areas on a smaller scale is the so-called Multi-Voxel-Pattern-Analysis (MVPA, Schwarzkopf and Rees 2011). In simplified terms, this analysis tries to determine those voxels that either accurately predict the outcome of a behavioral task carried out during scanning (e.g. intelligibility rating, Okada et al. 2010) or accurately distinguish between stimulus properties (e.g. place, manner, and voicing, Arsenault and Buchsbaum 2015).

3.1 Phonology in the temporal lobes

As illustrated in Table 1, the area most commonly found in tasks supposedly tapping into phonological processing is the superior temporal sulcus (STS) in the left and right hemisphere (with an inconsistent left dominance that shall not be further discussed here, for further discussion, see e.g. Wolmetz, Poeppel, and Rapp 2011). Afforded to limits of spatial resolution and difficulties in identifying sulci (relative depressions in the cortex), activations are also reported from superior temporal gyrus (located dorsally to STS) as well as middle temporal gyrus (located ventrally to STS, see Figure 2).

While there seems to be agreement on the role of STS for phonological and phonetic processing, disagreement exists with regard to the functional roles of its subdivisions into anterior, middle, and posterior parts. Notwithstanding some variability in labeling these subdivisions, some researchers claim that posterior STS shows the highest degree of acoustic invariance, and thus the highest degree of phonological abstraction (Okada et al. 2010). The authors propose a hierarchical progression from Heschl's gyrus through anterior STS to posterior STS along which acoustic invariance increases. By contrast, Liebenthal et al. (2010) propose that posterior STS houses short-term representations for relevant sound features, with a relatively low level of abstraction from acoustic details. Scott et al. (2000) showed that posterior STS is sensitive to phonetic features, irrespective of intelligibility, while the sensitivity of anterior STS to phonetic features crucially depends on intelligibility. Obleser et al. (2006) provided evidence that anterior STS enables distinctions across phonological/phonetic feature categories, such as tongue height (contrasting mid with high vowels) and place of articulation (contrasting coronal with dorsal vowels). Deschamps, Baum, and Gracco (2015) suggest that anterior-to-mid portions of STS are particularly sensitive to phonemic and syllabic information, while mid-to-posterior portions are more sensitive to long-term phonological (i.e. lexical) information. Okada et al. (2010) argue for a gradient of abstraction from anterior STS (with more reliance on spectro-temporal properties) to posterior STS (with more reliance on abstract, categorical properties). This gradient of abstraction seems somewhat at odds with the anatomical configuration of the temporal lobes and the gradient of abstraction originating from Heschl's gyrus, as illustrated by Obleser and Eisner (2009) and discussed in Humphries et al. (2014). Spatially, the anterior part of STS lies in vicinity of the anterior temporal lobe (ATL), an area that has been attributed to support lexical-semantic processing (Westerlund and Pylkkänen 2014). Thus, from a conceptual point of view, it seems more likely that abstract (and meaning-distinguishing) phonological features are supported by anterior, rather than posterior STS. This assumption is in-line with the sensitivity of anterior STS to phonetic feature in intelligible as compared to unintelligible stimuli (Scott et al. 2000). Intelligibility – if not solely referring to perceived clarity – describes the fact that acoustic sounds can be attributed meaning. Finally, the progression of information from HG to anterior STS follows the ventral stream of auditory processing (DeWitt and Rauschecker 2012; Hickok 2012; Poeppel, Idsardi, and van Wassenhove 2008; Rauschecker 2012), realized as white-matter fiber tracts (Friederici 2015). Following the assumptions of Hickok and Poeppel (2007), the ventral stream describes the interface of the phonological network with lexical information and the combinatorial network and thus stands for a cortical implementation of the sound-meaning relation of which the smallest unit is the distinctive (phonological) feature. Connectivity analyses of the STS confirm its differing involvement in the ventral and dorsal stream. Erickson, Rauschecker, and Turkeltaub (2017) provide evidence that pSTS connects to the dorsal stream and areas in IPL, precentral gyrus and supplementary motor area, while aSTS is linked to the ventral stream and to areas in the anterior temporal lobe and in the IFG. Finally, a recent study showed that mid-STS supports modality-independent aspects of phonological processing (Oron et al. 2016). A similar region (anterior-to-mid-STS) was found to be sensitive to phonological features oppositions (Scharinger et al. 2016), following the hypotheses of Lahiri and colleagues (Lahiri & Reetz 2002, 2010) regarding asymmetric activations.

3.2 Phonology outside the temporal lobes

Outside the temporal lobes, phonological processing seems additionally to be supported by parietal and frontal areas, in particular, inferior parietal lobe (IPL) and neighboring supramarginal gyrus (SMG), as well as inferior frontal gyrus (IFG, see Figure 2).

The role of the IFG for phonological processing appears to be related to the processing of sequential aspects of speech, as for instance required during segmentation (e.g. necessary during same-different tasks on initial speech sounds in non-rhyming words) and sequencing processes during perception (Burton 2001; Gelfand and Bookheimer 2003; LoCasto et al. 2004). Note that speech segmentation is generally assumed to be guided by phonological structure (for a review, see Boll-Avetisyan 2018, this volume). The involvement of IFG thus seems to depend both on the nature of the task and on the nature of the stimulus: As can be discerned from Table 1, studies with word and pseudoword stimuli are much more likely to report IFG involvement than studies with phoneme and syllable stimuli. When IFG was found in studies with phoneme or syllable stimuli, they most likely used a task with sequential nature (e.g. Gelfand and Bookheimer 2003). The importance of IFG for processing sequential or combinatorial aspects in morphophonologically complex words (as found in English regular past tense) is further highlighted in Joanisse and Seidenberg (2005). In sum, the involvement of IFG in phonological processing seems to become increasingly important with increasing stimulus complexity and decreasing distance to processing levels at the interface of phonology and syntax (Meyer et al. 2012). Furthermore, the study by Joanisse and Seidenberg (2005) also illustrates how phonology interacts with morphology and that this interaction may be particularly supported by the IFG. This also relates to the observation that morphological operations can be constrained by phonological (prosodic) properties of words (Domahs, Domahs, and Kauschke 2018, this volume).

Parietal areas, on the other hand, appear particularly relevant for tasks with a (working) memory component (e.g. IPL in Harinen and Rinne 2013). IPL and SMG have been ascribed important function in the articulatory (Baddeley, Lewis, and Vallar 1984) or phonological loop (Aboitiz, Aboitiz, and García 2010), a specialized sensorimotor circuit connecting temporal areas with parietal and frontal regions. The phonological loop bears close resemblance to the proposed dorsal stream, connecting a sensorimotor interface at the border of posterior temporal and parietal regions with the articulatory network in the IFG and premotor cortex (PMC). This dorsal stream complements a ventral stream, connecting a lexical interface in posterior parts of the MTG with a combinatorial network in anterior parts of the MTG (sound-to-meaning stream, cf. Hickok and Poeppel 2007).

Regarding phonological processing, the dorsal stream seems to aid categorical perception (Chevillet et al. 2013), particularly in situations of sensory ambiguity, when e.g. phonemes are ambiguous on a phonetic continuum (for fMRI and brain stimulation evidence, see Callan et al. 2010; D'Ausilio et al. 2012; Möttönen, van de Ven, and Watkins 2014). These findings suggest that articulatory information is not part of the auditory-based phonological codes in STS, but may aid to enrich these codes with categorical information, particularly in difficult listening situations.

3.3 Notes of caution

When evaluating imaging experiments on phonological processing, two important aspects have to be considered: First, one should be aware of the methodological limits and the types of contrasts underlying the identification of relevant brain regions. Second, it is very important to consider to what extent the nature of the stimuli and the properties of the task (if any) actually tap into phonological, rather than phonetic or acoustic processing (see Poeppel 1996 for more details on these and further points with respect to an imaging technique preceding fMRI), and whether processing allows for conclusions about the localization of phonological representations.

Regarding the first point, a critical evaluation of imaging studies needs to consider whether the contrast between the experimental task and the control task allows for a meaningful functional interpretation of the emerging brain regions from the contrast analysis. For instance, Callan et al. (2004) employ a continuous scanning design, and include silent trials as baseline. However, this effectively means that they compare activations resulting from scanner noise with activations resulting from scanner noise and additional syllabic material. Thus, brain areas emerging from this contrast indicate the additional need of neural resources in processing (recognizing) syllables on the background of scanner noise, but do not directly establish a phonological function to the identified areas. Such phonological function must then be induced from independent evidence (see Caplan 2009 for an in-depth discussion of this topic).

Similarly, if not even more critical, is the second point: Attributing functional relevance to a certain brain region requires a precise definition of this function, and ideally an embedding in a functional architecture, that specifies how computations occurring at various stages of speech processing are ordered relative to one another. For phonological processing, it must be clearly stated what computations operate on which representations, and how these operations are ordered and feed from one level to the next, or back to a previous one. In most studies, this accuracy in description is lacking. The majority of the studies illustrated in Table 1 furthermore reduce phonological processing to the processing of phonemic stimuli, in contrast to non-phonemic or unintelligible stimuli. Only some studies more explicitly examine properties beyond the phonemic status of the experimental stimuli, for instance, sonority in onset clusters (Deschamps, Baum, and Gracco 2015), phonotactic licensing (Jacquemot et al.

2003; Obrig, Mentzel, and Rossi 2016), or phonological neighborhood density (Okada and Hickok 2006). Even the study by Lawyer and Corina (2014), suggesting separable areas in STS that selectively respond to place of articulation and voicing distinctions, needs to be complemented by cross-linguistic investigation as e.g. provided by Callan et al. (2004) in order to speak to language-specific (phonological) interpretations of these phonetic classifications.

On a technical note, functional brain imaging studies on phonological processing are almost exclusively based on activations, rather than deactivations, of certain brain areas (compared to baselines). However, as mentioned earlier, deactivations can also be attributed functional relevance (e.g. Allison et al. 2000), such that future research should become less resilient in ignoring such findings.

Finally, as mentioned above, localization of phonological processing areas does not necessarily mean that these areas house a phonological representation. It is rather the case that they provide the neural resources to operate on aspects of phonological representations. As has been shown before, phonological representations in the brain are rather distributed than focal (e.g. Arsenault and Buchsbaum 2015). In this regard, bilateral STS provides neural resources to operate on auditory abstract aspects of phonological representations, of which the phonological feature is a prominent candidate example. As will be exemplified in the next section, ERPs provide an ideal measure to further examine the feature-based aspects of phonological representations.

4 Electrophysiological investigations of phonological representations

As illustrated in Figure 1, electrophysiological measures (i.e. brain electric and magnetic responses) offer superior temporal resolution compared to the hemodynamic response and are therefore very suitable to examine the time course phonological representations are operated on. Similar to fMRI investigations, electrophysiological studies try to uncover processing mechanisms that are distinct from acoustic processing. To that end, electrophysiological studies on phonology employ an acoustic control condition or acoustically controlled designs that enable a clear attribution of an observed effect to aspects of phonological processing or phonological representations.

Electrophysiological measures on phonology are commonly based on so-called Event-Related Potentials (ERPs), that is, time-locked responses to sensory stimuli with distinct negative and positive deflections and topographies that have received functional interpretations. Regarding phonological processing and phonological representations, a candidate deflection is the Mismatch Negativity (MMN). The MMN is an automatic, pre-attentive change detection response of the brain and an index of auditory regularity violations (Näätänen et al. 2007; Schröger 2005). Most commonly, the MMN is seen as difference between the response to infrequently occurring deviant stimuli, and the response to frequently occurring standard stimuli presented in a (passive) oddball paradigm (see Figure 3). In this paradigm, the frequently and repeatedly occurring standard stimulus is assumed to generate rules, predictions, or abstract stimulus traces that are subsequently violated by the rarely and unexpectedly occurring deviant stimulus. This deviant stimulus differs in one or more aspects from the standard stimulus. The MMN as a difference wave form peaks between 120 and 250 ms after deviance onset and has sources in the temporal plane as well as in frontal cortices (Doeller et al. 2003). A recent fMRI study has provided evidence for MMN sources in STS, therefore strengthening a possible phonological interpretation of the MMN response (Shtyrov, Osswald, and Pulvermüller 2008).

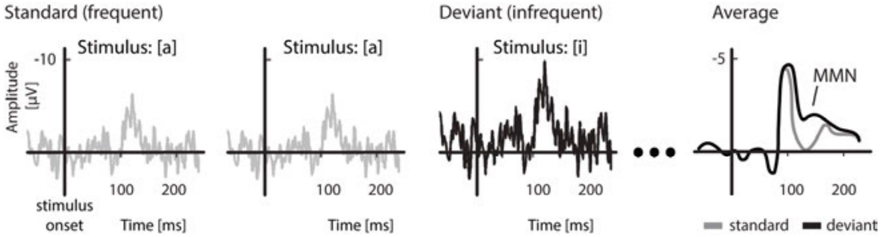


Fig. 3: Illustration of Mismatch Negativity (MMN) response in an oddball paradigm. Often-repeated standard stimuli (here: vowel [a]) elicit an Event-Related Potential (ERP) that differs from that to rarely occurring deviant stimuli (here: vowel [i]). The difference is most pronounced between 120 and 250 ms after deviance onset. Single-trial electrophysiological responses are shown on the left, while averaged responses are presented in the right-most panel.

While the magnitude of the MMN response primarily depends on the physical difference between standard and deviant stimuli, the suitability of the MMN for investigation phonological processing is based on the observation that the phonemic status of sounds in an oddball paradigm additionally enhances the MMN response (Dehaene-Lambertz, Dupoux, and Gout 2000; Kazanina, Phillips, and Idsardi 2006; Näätänen et al. 1997). The observation that the magnitude of the

MMN is not solely dependent on the acoustic difference between standard and deviant stimuli (directly compared in e.g. Winkler et al. 1999) made the response a candidate index of language- and speech-specific processing at early latencies. Apart from tapping into phonemic properties, the usefulness of the MMN for examining phonological processes and representations rests on the observation that it indexes the phonotactic probability of particular sound sequences (Bonte et al. 2005), phonological rules such as final devoicing in German (Truckenbrodt et al. 2014), violations of vowel harmony (Aaltonen et al. 2008; Scharinger, Idsardi, and Poe 2011), dialect-specific tonal properties (Werth et al. 2018, this volume) and allophonic relations of speech sounds (Miglietta, Grimaldi, and Calabrese 2013).

In general, MMN studies try to constrain differences between standard and deviant stimuli to the contrasts of interest and usually investigate the electrophysiological response to the same acoustic stimulus in different positions within the oddball paradigm: Thus, the MMN is obtained by subtracting the ERP of a stimulus in standard position from the ERP of the acoustically identical stimulus in deviant position (identity MMN, Pulvermüller et al. 2006). Despite such acoustic control, MMN research has found an effect of asymmetry in that the magnitude of the MMN depended on the directionality of the standard-deviant reversal. For instance, Maiste et al. (1995) found that the MMN based on the deviant [da] preceded by the standard [ga] was larger than the MMN based on the deviant [ga], preceded by the standard [da]. At first sight, this asymmetry appears to result from acoustic properties. Studies explicitly examining the direction of changes in the frequency domain (Jacobsen and Schröger, 2001; Peter, McArthur, and Thompson 2010) are suggestive in this regard, but fail to explain results as, for instance, reported in Scharinger et al. (2012). One interpretation from proponents of phonological underspecification is that these asymmetries reflect differences in phonological specifications. For instance, Eulitz and Lahiri (2004) assumed that a single coronal vowel such as German [ø] would activate a memory trace without a specification for place of articulation when presented in standard position of an oddball paradigm. Because of this place of articulation underspecification, a non-coronal deviant vowel [o], albeit an acoustic deviant, does not mismatch with the representation of [ø], and should therefore only elicit an acoustically-based MMN. In the reverse case, a specified non-coronal (dorsal) vowel [o] in standard position should tap into a more specific memory trace. A deviant vowel with a different place of articulation would then provide not only an acoustic, but also a featural mismatch, resulting in an enhanced MMN response. This pattern of responses was indeed found in the study of Eulitz and Lahiri (2004).

Tab. 2: Overview of studies reporting asymmetric MMN effects with speech stimuli

Article	Stimulus Type	Stimulus oppositions	Technique (Channels)	Listeners	Effect	Phenomenon
Cornell, Lahiri & Eulitz (2013)	German pseudowords	[eni] – [ezi]; [edi] – [egi]; [eni] – [edi]; reversal	EEG (64)	German	asymmetry: larger MMN for less specific deviant	phonological under-specification
Cornell, Lahiri & Eulitz (2011)	German words, pseudowords	[e:] – [ø:]; [o:] – [ø:]; reversal (in word/pseudoword)	EEG (64)	German	asymmetry: larger MMN for less specific deviant	phonological under-specification
Eulitz & Lahiri (2004)	German vowels	[e:] – [ø:]; [o:] – [ø:]; [e:] – [o:]; reversal	EEG (65)	German	asymmetry: larger MMN for less specific deviant	phonological under-specification
Hestvik & Durvasula (2015)	English syllables	[ta] – [da]; reversal	EEG (128)	English	asymmetry: larger MMN for less specific deviant	phonetic vs. phonological specification
Politzer-Ahles et al. (2016)	Mandarin syllables	Tone3 – Tone 4; Tone 3 – Tone 2; reversal	EEG (66)	Mandarin, non-Mandarin	asymmetry: larger MMN for less specific deviant	perceptual vs. phonological under-specification
Scharinger et al. (2012)	German words	[laks] – [lats]; reversal	EEG (64)	German	asymmetry: larger MMN for less specific deviant	phonological under-specification
Scharinger et al. (2011)	English syllables	[aja] – [awa]; [aʒa] – [ava]; reversal	MEG (157)	English	larger MMN for bilabials	articulatory specificity
Scharinger, Idsardi & Poe (2011)	Turkish vowels	[ɯ] – [œ]; [ɛ] – [œ]; reversal	MEG (157)	Turkish	asymmetry: larger MMN for VH violation	vowel harmony

Article	Stimulus Type	Stimulus oppositions	Technique (Channels)	Listeners	Effect	Phenomenon
Scharinger, Monahan & Idsardi (2012)	English vowels	[ɛ] – [æ]; [ɪ] – [æ]; reversal	MEG (157)	English	asymmetry: larger MMN for less specific deviant	phonological underspecification
Scharinger, Monahan & Idsardi (2016)	English vowels	[ɛ] – [ɪ]; [ɪ] – [æ]; reversal	MEG (157)	English	asymmetry: larger MMN for less specific deviant	phonological underspecification
Schluter, Politzer-Ahles & Almeida (2016)	English fricatives	[s] – [h], [s] – [f]; reversal	EEG (34)	English	asymmetry: larger MMN for less specific deviant, dependent on feature geometry	phonological underspecification
Werth et al. (2017)	German words	[ʃa:^{2}l] – [ʃal^2]; reversal	EEG (26)	Standard-German, Moselle-Franconian	asymmetry: earlier tonal effect for Moselle-Franconian	dialect-dependent phonology

Asymmetries in MMN responses of the sort described above have been replicated in different languages (Scharinger, Idsardi, and Poe 2011; Scharinger, Monahan, and Idsardi 2012) and for different stimuli (Cornell, Lahiri, and Eulitz 2011, 2013; Scharinger et al. 2012; Scharinger, Eulitz, and Lahiri 2010). Notably, asymmetries were not constrained to MMN amplitudes, but also held for MMN (peak) latencies (e.g. Werth et al. 2018, this volume). The asymmetric patterns were either compatible with the underspecification approach as envisaged by Lahiri and Reetz (2010) or reflected (dialect-)specific phonological representations. Table 2 lists the most relevant studies whose interpretations are based on MMN asymmetries.

Importantly, it has been acknowledged in recent work (Scharinger et al. 2012; Scharinger, Monahan, and Idsardi 2012) that the phonological underspecification approach is compatible with the predictive coding approach in the auditory neurosciences (Friston 2005). In short, the predictive coding approach assumes that perception results from the integration of available sensory evidence with prior assumptions, gathered from diverse sources such as contextual stimulus statistics (rule extraction) or long-term memory (Bar 2009). The predictive coding approach has recently also been used to account for MMN generation (Baldeweg 2006; Garrido et al. 2009; Winkler 2007). The sequence of standard sounds generates inferences about forthcoming sound events, i.e., provides an anticipated continuation of the standard sequence (acoustic model). Importantly, the model can be more or less confident in inferring future sound events, and consequently will show larger mismatch responses if a highly confident inference is violated. Scharinger et al. (2012) suggested that inferences from more specific speech sound representations in standard position are stronger than inferences emerging from sparsely represented speech sounds in standard position.

For instance, a dorsal speech sound in standard position (vowel [o] or consonant [k]) generates strong inferences regarding the place of articulation of forthcoming sounds. Encountering a coronal deviant (vowel [ø] or consonant [t]) results in a strong prediction error, and a large MMN response is generated. In the reverse case, a coronal speech sound in standard position (vowel [ø] or consonant [t]) generates a rather weak inference regarding the place of articulation of an upcoming sound. Consequently, a deviant sound (vowel [o] or consonant [k]) provides a weaker prediction error, and, in turn, elicits a reduced MMN.

In general, the predictive coding approach to speech and language processing has become a fruitful enterprise. It provides unifying descriptions of cortical and subcortical mechanisms and pays tribute to the observation that speech and language are inherently predictive in nature (Tavano and Scharinger 2015). Furthermore, the emphasis on influences from prior (long-term memory) information on sensory perception is well-suited to describe cross-linguistic differences in interpreting the same acoustic stimulus, depending on one's native phonology.

4.1 Notes of caution and perspectives

As with fMRI experiments, it is important to consider potential shortcomings of MMN experiments on phonological representations and phonological processing. Critically, MMN experiments on phonology must provide good ways of excluding acoustic confounds. Commonly, this is achieved by looking at the identity MMN. While recent studies take this as standard procedure, some older studies are based on the standard-deviant difference between acoustically different stimuli (e.g. Aaltonen et al. 2008). A further crucial point is the absence of a significant difference between standard and deviant responses in a particular condition (e.g. Bomba, Choly, and Pang 2011), particularly if statistical analyses are based on the difference wave forms (rather than on separate standard and deviant ERPs).

MMN studies with speech material need to strictly control for confounding factors such as frequency of occurrence (Alexandrov et al. 2011) or phonotactic probability (Bonte et al. 2005) that may influence the MMN response. Another confounding factor – particularly for underspecification interpretations – concerns speech sounds whose representations may be enriched by articulatory or visual information. For instance, Scharinger et al. (2011) found consistently larger MMNs to deviants with labial consonants, contradicting the hypotheses of underspecification. Perhaps, labial speech sounds recruit additional processing areas such that the mismatch response is enhanced by additional MMN sources, thereby masking a possible underspecification effect.

Finally, oddball paradigms do certainly not present everyday listening conditions and provide a rather artificial situation. Nevertheless, recent findings suggest that ERP components at MMN latencies and with MMN-like topographies are not restricted to oddball paradigms, but can also be elicited by more natural linguistic material, such as spoken sentences (Bendixen et al. 2014). In this regard, a latter negativity with latencies between 200 and 400 ms post stimulus onset has been identified as an index of phonological processing. This so-called Phonological Mismatch Negativity (PMN; Connolly and Phillips 1994; Praamstra, Meyer, and Levelt 1994), later re-named as Phonological Mapping Negativity (Newman and Connolly 2009) resembles a semantic N400 (negativity at around 400 ms) in that its amplitude scales with the expectancy of a particular sound in either syllable onset or offset. In some studies, the response shows clearer separation from the semantic N400 (van den Brink, Brown, and Hagoort 2001) or from phonologically more sensitive components (Friedrich, Lahiri, and Eulitz 2008; Friedrich, Schild, and Röder 2009) such that its phonological character seems to be restricted to phoneme-based expectancies during speech perception. While a more detailed discussion of the PMN is unfortunately beyond the scope of this review, its sensitivity to speech sound expectancies is clearly beneficial for future studies that try to bring together phonological underspecification with predictive coding.

5 Conclusions

Empirical approaches to phonology that focus on the neurobiological bases of phonological representations have provided insights into possible units that phonological processes operate on. These units appear to be rather invariant and may be ideally conceived of as phonological features that are based on acoustic properties for which auditory neurons seem to have specific preferences (Chang et al. 2010). This review illustrated two main techniques of brain imaging – fMRI and EEG – and discussed to what extent these techniques can further our understanding of how specialized regions of the human brain support phonological processing. While none of the studies can claim to have uncovered the phonological representation in its entirety, the overview provided here suggests that phonological representations have a distributed nature and cannot be restricted to one processing area. However, there is converging evidence that the STS is a key region for sensitivity to aspects of phonological representations that can be described by phonological features. Such an approach is compatible with recent findings of hierarchical processing within auditory regions (Humphries et al. 2014) and with the assumption of predictive coding (Friston 2005). The latter point is further illustrated with findings from ERP studies, the pattern of which are compatible with both a phonological underspecification account and a predictive coding framework.

References

Aaltonen, Olli, Ake Hellström, Maija S. Peltola, Janne Savela, Henna Tamminen & Heidi Lehtola. 2008. Brain responses reveal hardwired detection of native-language rule violations. Neuroscience Letters 444(1). 56–59.

Aboitiz, Francisco, Sebastián Aboitiz & Ricardo R. García. 2010. The Phonological Loop: A key innovation in human evolution. Current Anthropology 51(S1). S55–S65.

Alexandrov, Alexander A., Daria O. Boricheva, Friedemann Pulvermüller & Yury Shtyrov. 2011. Strength of word-specific neural memory traces assessed electrophysiologically. PLoS ONE 6(8). e22999.

Allison, J. D., K. J. Meador, D. W. Loring, R. E. Figueroa & J. C. Wright. 2000. Functional MRI cerebral activation and deactivation during finger movement. Neurology 54(1). 135–142.

Arsenault, Jessica S. & Bradley R. Buchsbaum. 2015. Distributed neural representations of phonological features during speech perception. The Journal of Neuroscience 35(2). 634–642.

Ashtari, Manzar, Todd Lencz, Paula Zuffante, Robert Bilder, Tana Clarke, Alan Diamond, John Kane & Philip Szeszko. 2004. Left middle temporal gyrus activation during a phonemic discrimination task. Neuroreport 15(3). 389–393.

Baddeley, Alan, V. Lewis & G. Vallar. 1984. Exploring the articulatory loop. Quarterly Journal of Experimental Psychology 36. 233–252.

Baldeweg, T. 2006. Repetition effects to sounds: Evidence for predictive coding in the auditory system. Trends in Cognitive Sciences 10(3). 93–94.

Bar, Moshe. 2009. The proactive brain: Memory for predictions. Philosophical Transactions of the Royal Society of London B: Biological Sciences 364(1521). 1235–1243.

Bendixen, Alexandra, Mathias Scharinger, Antje Strauss & Jonas Obleser. 2014. Prediction in the service of speech comprehension: Modulated early brain responses to omitted speech segments. Cortex 53. 9–26.

Binder, Jeffrey R., J. A. Frost, T. A. Hammeke, P. S. Bellgowan, J. A. Springer, J. N. Kaufman & E. T. Possing. 2000. Human temporal lobe activation by speech and nonspeech sounds. Cerebral Cortex 10(5). 512–528.

Boll-Avetisyan, Natalie. 2018. The role of phonological structure in speech segmentation by infants and adults: a review and methodological considerations. In Christiane Ulbrich, Alexander Werth & Richard Wiese (eds.), Empirical approaches to the phonological structure of words (Linguistische Arbeiten 567), 213–240. Berlin & Boston: De Gruyter.

Bomba, Marie D., David Choly & Elizabeth W. Pang. 2011. Phoneme discrimination and mismatch negativity in English and Japanese speakers. NeuroReport 22(10). 479–483.

Bonte, Milene L., Holger Mitterer, Nadia Zellagui, Hanne Poelmans & Leo Blomert. 2005. Auditory cortical tuning to statistical regularities in phonology. Clinical Neurophysiology 116. 2765–2774.

Brink, Daniëlle van den, Colin M. Brown & Peter Hagoort. 2001. Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. Journal of Cognitive Neuroscience 13(7). 967–985.

Buchsbaum, Bradley R., Gregory Hickok & Colin Humphries. 2001. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. Cognitive Science 25(5). 663–678.

Burton, Martha W. 2001. The role of inferior frontal cortex in phonological processing. Cognitive Science 25(5). 695–709.

Burton, Martha W., Paul LoCasto, Donna Krebs-Noble & Rao P. Gullapalli. 2005. A systematic investigation of the functional neuroanatomy of auditory and visual phonological processing. NeuroImage 26(3). 647–661.

Burton, Martha W., Steven L. Small & Sheila E. Blumstein. 2000. The role of segmentation in phonological processing: An fMRI investigation. Journal of Cognitive Neuroscience 12(4). 679–690.

Buxton, Richard B. 2009. Introduction to Functional Magnetic Resonance Imaging: Principles and Techniques. New York: Cambridge University Press.
Callan, Daniel, Akiko Callan, Mario Gamez, Masa-aki Sato & Mitsuo Kawato. 2010. Premotor cortex mediates perceptual performance. NeuroImage 51(2). 844–858.
Callan, Daniel E., Jeffery A. Jones, Akiko M. Callan & Reiko Akahane-Yamada. 2004. Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory–auditory/orosensory internal models. NeuroImage 22(3). 1182–1194.
Caplan, David. 2009. Experimental design and interpretation of functional neuroimaging studies of cognitive processes. Human Brain Mapping 30(1). 59–77.
Carbonell, Kathy M. & Andrew J. Lotto. 2014. Speech is not special… again. Language Sciences 5. 427.
Carlson, Nicole L., Vivienne L. Ming & Michael Robert DeWeese. 2012. Sparse codes for speech predict spectrotemporal receptive fields in the inferior colliculus. PLoS Computational Biology 8(7). e1002594.
Chang, Edward F., Jochem W. Rieger, Keith Johnson, Mitchel S. Berger, Nicholas M. Barbaro & Robert T. Knight. 2010. Categorical speech representation in human superior temporal gyrus. Nature Neuroscience 13(11). 1428–1432.
Chevillet, Mark A., Xiong Jiang, Josef P. Rauschecker & Maximilian Riesenhuber. 2013. Automatic phoneme category selectivity in the dorsal auditory stream. The Journal of Neuroscience 33(12). 5208–5215.
Chomsky, Noam & Morris Halle. 1968. The Sound Pattern of English. New York: Harper and Row.
Connolly, J. F. & N. A. Phillips. 1994. Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. Journal of Cognitive Neuroscience 6(3). 256–266.
Cornell, Sonia A., Aditi Lahiri & Carsten Eulitz. 2011. "What you encode is not necessarily what you store": Evidence for sparse feature representations from mismatch negativity. Brain Research 1394. 79–89.
Cornell, Sonia A., Aditi Lahiri & Carsten Eulitz. 2013. Inequality across consonantal contrasts in speech perception: Evidence from mismatch negativity. Journal of experimental psychology: Human perception and performance 39(3). 757–772.
D'Ausilio, Alessandro, Ilaria Bufalari, Paola Salmas & Luciano Fadiga. 2012. The role of the motor system in discriminating normal and degraded speech sounds. Cortex 48(7). 882–887.
Davis, Matthew H. & Ingrid S. Johnsrude. 2003. Hierarchical processing in spoken language comprehension. The Journal of Neuroscience 23(8). 3423–3431.
Davis, Matthew H. & Ingrid S. Johnsrude. 2007. Hearing speech sounds: top-down influences on the interface between audition and speech perception. Hearing research 229(1–2). 132–147.
Dehaene-Lambertz, Ghislaine, Emmanuel Dupoux & Ariel Gout. 2000. Electrophysiological correlates of phonological processing: A cross-linguistic study. Journal of Cognitive Neuroscience 12(4). 635–647.
Dehaene-Lambertz, Ghislaine, Christophe Pallier, Willy Serniclaes, Liliane Sprenger-Charolles, Antoinette Jobert & Stanislas Dehaene. 2005. Neural correlates of switching from auditory to speech perception. NeuroImage 24(1). 21–33.

Desai, R., E. Liebenthal, E. Waldron & J. R. Binder. 2008. Left posterior temporal regions are sensitive to auditory categorization. Journal of Cognitive Neuroscience 20(7). 1174–1188.
Deschamps, Isabelle, Shari R. Baum & Vincent L. Gracco. 2015. Phonological processing in speech perception: What do sonority differences tell us? Brain and Language 149. 77–83.
DeWitt, Iain & Josef P. Rauschecker. 2012. Phoneme and word recognition in the auditory ventral stream. Proceedings of the National Academy of Sciences of the United States of America 109(8). E505–514.
DeWitt, Iain & Josef P. Rauschecker. 2016. Convergent evidence for the causal involvement of anterior superior temporal gyrus in auditory single-word comprehension. Cortex 77. 164–166.
Doeller, Christian F., Bertram Opitz, Axel Mecklinger, Christoph Krick, Wolfgang Reith & Erich Schröger. 2003. Prefrontal cortex involvement in preattentive auditory deviance detection: neuroimaging and electrophysiological evidence. NeuroImage 20(2). 1270–1282.
Domahs, Ulrike, Frank Domahs & Christina Kauschke. 2018. The morphology-prosody interface in typically developing and language-impaired populations: An overview and insights from experimental approaches. In Christiane Ulbrich, Alexander Werth & Richard Wiese (eds.), Empirical approaches to the phonological structure of words (Linguistische Arbeiten 567), 95–120. Berlin & Boston: De Gruyter.
Erickson, Laura C., Josef P. Rauschecker & Peter E. Turkeltaub. 2017. Meta-analytic connectivity modeling of the human superior temporal sulcus. Brain Structure and Function 222(1). 267–285.
Eulitz, Carsten & Aditi Lahiri. 2004. Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. Journal of Cognitive Neuroscience 16(4). 577–583.
Fant, Gunnar. 1960. Acoustic Theory of Speech Production. The Hague: Mouton de Gruyter.
Friederici, Angela D. 2012. The cortical language circuit: From auditory perception to sentence comprehension. Trends in Cognitive Sciences 16(5). 262–268.
Friederici, Angela D. 2015. White-matter pathways for speech and language processing. Handbook of Clinical Neurology 129. 177–186.
Friedrich, Claudia, Aditi Lahiri & Carsten Eulitz. 2008. Neurophysiological evidence for underspecified lexical representations: Asymmetries with word initial variations. Journal of Experimental Psychology: Human Perception and Performance 34(6). 1545–1559.
Friedrich, Claudia, Ulrike Schild & Brigitte Röder. 2009. Electrophysiological indices of word fragment priming allow characterizing neural stages of speech recognition. Biological Psychology 80. 105–113.
Friston, K. 2005. A theory of cortical responses. Philosophical Transactions of the Royal Society of London B: Biological Sciences 360(1456). 815–815.
Friston, Karl. 2010. The free-energy principle: a unified brain theory? Nature Reviews Neuroscience 11(2). 127–138.
Garrido, Marta I., James M. Kilner, Klass E. Stephan & Karl J. Friston. 2009. The mismatch negativity: A review of underlying mechanisms. Clinical Neurophysiology 120. 453–463.
Gelfand, Jenna R. & Susan Y. Bookheimer. 2003. Dissociating neural mechanisms of temporal sequencing and processing phonemes. Neuron 38(5). 831–842.
Gladziejewski, Pawel. 2015. Predictive coding and representationalism. Synthese 193(2). 559–582. DOI: 10.1007/s11229-015-0762-9.
Hansen, Peter C., Morten L. Kringelbach & Riitta Salmelin. 2010. MEG: An Introduction to Methods. New York: Oxford University Press.

Harinen, Kirsi & Teemu Rinne. 2013. Activations of human auditory cortex to phonemic and nonphonemic vowels during discrimination and memory tasks. NeuroImage 77. 279–287.
Hestvik, Arild & Karthik Durvasula. 2015. Neurobiological evidence for voicing underspecification in English. Brain and Language 152. 28–43.
Hickok, Gregory. 2012. The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. Journal of Communication Disorders 45(6). 393–402.
Hickok, Gregory & David Poeppel. 2007. The cortical organization of speech processing. Nature Reviews Neuroscience 8(5). 393–402.
Humphries, Colin, Merav Sabri, Kimberly Lewis & Einat Liebenthal. 2014. Hierarchical organization of speech perception in human auditory cortex. Frontiers in Neuroscience 8. 406.
Husain, Fatima T., Stephen J. Fromm, Randall H. Pursley, Lara A. Hosey, Allen R. Braun & Barry Horwitz. 2006. Neural bases of categorization of simple speech and nonspeech sounds. Human Brain Mapping 27(8). 636–651.
Jacobsen, T. & E. Schröger. 2001. Is there pre-attentive memory-based comparison of pitch? Psychophysiology 38(4). 723–727.
Jacquemot, Charlotte, Christophe Pallier, Denis LeBihan, Stanislas Dehaene & Emmanuel Dupoux. 2003. Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. The Journal of Neuroscience 23(29). 9541–9546.
Jakobson, Roman. 1939. Observations sur le classement phonologique des consonnes. In Edgard Blancquaert & Willem Pée (eds.), Proceedings of the 3rd International Conference of Phonetic Sciences, 34–41. Ghent.
Joanisse, Marc F. & Mark S. Seidenberg. 2005. Imaging the past: neural activation in frontal and temporal regions during regular and irregular past-tense processing. Cognitive, affective & behavioral neuroscience 5(3). 282–296.
Kazanina, Nina, Colin Phillips & William Idsardi. 2006. The influence of meaning on the perception of speech sounds. Proceedings of the National Academy of Sciences of the United States of America 103(30). 11381–11386.
Kutas, Marta, Cyma van Petten & Robert Kluender. 2006. Psycholinguistics Electrified II (1994–2005). In Matthew Taxler & Morton Ann Gernsbacher (eds.), Handbook of Psycholinguistics, 659–724. New York: Academic Press.
Kwong, K. K., J. W. Belliveau, D. A. Chesler, I. E. Goldberg, R. M. Weisskoff, B. P. Poncelet, D. N. Kennedy, B. E. Hoppel, M. S. Cohen & R. Turner. 1992. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proceedings of the National Academy of Sciences of the United States of America 89(12). 5675–5679.
Lahiri, Aditi, Letitia Gewirth & Sheila E. Blumstein. 1984. A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: evidence from a cross-language study. Journal of the Acoustical Society of America 76(2). 391–404.
Lahiri, Aditi & William Marslen-Wilson. 1991. The mental representation of lexical form: A phonological approach to the recognition lexicon. Cognition 38. 245–294.
Lahiri, Aditi & Henning Reetz. 2002. Underspecified recognition. In Carlos Gussenhoven & Natasha Warner (eds.), Laboratory Phonology VII, 637–677. Berlin: Mouton de Gruyter.
Lahiri, Aditi & Henning Reetz. 2010. Distinctive features: Phonological underspecification in representation and processing. Journal of Phonetics 38. 44–59.
Lawyer, Laurel & David Corina. 2014. An investigation of place and voice features using fMRI-adaptation. Journal of Neurolinguistics 27(1).

Lee, Jin Hyung, Remy Durand, Viviana Gradinaru, Feng Zhang, Inbal Goshen, Dae-Shik Kim, Lief E. Fenno, Charu Ramakrishnan & Karl Deisseroth. 2010. Global and local fMRI signals driven by neurons defined optogenetically by type and wiring. Nature 465(7299). 788–792.

Liebenthal, Einat, Jeffrey R. Binder, Stephanie M. Spitzer, Edward T. Possing & David A. Medler. 2005. Neural substrates of phonemic perception. Cerebral Cortex 15(10). 1621–1631.

Liebenthal, Einat, Rutvik Desai, Michael M. Ellingson, Brinda Ramachandran, Anjali Desai & Jeffrey R. Binder. 2010. Specialization along the left superior temporal sulcus for auditory categorization. Cerebral Cortex 20(12). 2958–2970.

Liebenthal, Einat, Merav Sabri, Scott A. Beardsley, Jain Mangalathu-Arumana & Anjali Desai. 2013. Neural dynamics of phonological processing in the dorsal auditory stream. The Journal of Neuroscience 33(39). 15414–15424.

LoCasto, Paul C., Donna Krebs-Noble, Rao P. Gullapalli & Martha W. Burton. 2004. An fMRI investigation of speech and tone segmentation. Journal of Cognitive Neuroscience 16(9). 1612–1624.

Logothetis, Nikos K. & Brian A. Wandell. 2004. Interpreting the BOLD signal. Annual Review of Physiology 66. 735–769.

Lombardi, Linda. 1996. Postlexical rules and the status of privative features. Phonology 13. 1–38.

Maiste, A. C., A. S. Wiens, M. J. Hunt, M. Scherg & T. W. Picton. 1995. Event-related potentials and the categorical perception of speech sounds. Ear and Hearing 16(1). 68–90.

McCarthy, John J. & A. Taub. 1992. Review of: "Paradis, Carole & Jean-François Prunet (eds.). 1991. The special status of coronals: internal and external evidence (Phonetics and Phonology 2). San Diego: Academic Press". Phonology 9. 363–370.

McGettigan, Carolyn, Andrew Faulkner, Irene Altarelli, Jonas Obleser, Harriet Baverstock & Sophie K. Scott. 2012. Speech comprehension aided by multiple modalities: Behavioural and neural interactions. Neuropsychologia 50(5). 762–776.

McGettigan, Carolyn, Jane E. Warren, Frank Eisner, Chloe R. Marshall, Pradheep Shanmugalingam & Sophie K. Scott. 2010. Neural correlates of sublexical processing in phonological working memory. Journal of Cognitive Neuroscience 23(4). 961–977.

Meyer, Lars, Jonas Obleser, Alfred Anwander & Angela D. Friederici. 2012. Linking ordering in Broca's area to storage in left temporo-parietal regions: the case of sentence processing. NeuroImage 62(3). 1987–1998.

Miglietta, Sandra, Mirko Grimaldi & Andrea Calabrese. 2013. Conditioned allophony in speech perception: an ERP study. Brain and Language 126(3). 285–290.

Molfese, Dennis L., Victoria J. Molfese & Spencer Kelly. 2001. The use of brain electrophysiology techniques to study language: A basic guide for the beginning consumer of electrophysiology information. Learning Disability Quarterly 24(3). 177–188.

Möttönen, Riikka, Gido M. van de Ven & Kate E. Watkins. 2014. Attention fine-tunes auditory-motor processing of speech sounds. The Journal of Neuroscience 34(11). 4064–4069.

Näätänen, Risto, Anne Lehtokoski, Mietta Lennes, Marie Cheour, M. Huotilainen, Antti Ilvonen, Martti Vainio, Paavo Alku, R. J. Ilmoniemi, Aavo Luuk, Jueri Allik, Janne Sinkkonen & Kimmo Alho. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature 385. 432–434.

Näätänen, Risto, Petri Paavilainen, Teemu Rinne & Kimmo Alho. 2007. The mismatch negativity (MMN) in basic research of central auditory processing: A review. Clinical Neurophysiology 118(12). 2544–2590.

Newman, R. L. & J. F. Connolly. 2009. Electrophysiological markers of pre-lexical speech processing: Evidence for bottom–up and top–down effects on spoken word processing. Biological Psychology 80(1). 114–121.

Obleser, Jonas, Henning Boecker, Alexander Drzezga, Bernhard Haslinger, Andreas Hennenlotter, Michael Roettinger, Carsten Eulitz & Josef P. Rauschecker. 2006. Vowel sound extraction in anterior superior temporal cortex. Human Brain Mapping 27(7). 562–571.

Obleser, Jonas & Frank Eisner. 2009. Pre-lexical abstraction of speech in the auditory cortex. Trends in Cognitive Sciences 13(1). 14–19.

Obleser, Jonas, Frank Eisner & Sonja A. Kotz. 2008. Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. The Journal of Neuroscience 28(32). 8116–8123.

Obleser, Jonas & Sonja A. Kotz. 2010. Expectancy constraints in degraded speech modulate the language comprehension network. Cerebral Cortex 20(3). 633–640.

Obleser, Jonas, Jonas Zimmermann, John Van Meter & Josef P. Rauschecker. 2007. Multiple stages of auditory speech perception reflected in event-related FMRI. Cerebral Cortex 17(10). 2251–2257.

Obrig, Hellmuth, Julia Mentzel & Sonja Rossi. 2016. Universal and language-specific sublexical cues in speech perception: a novel electroencephalography-lesion approach. Brain 139(6). 1800–1816.

Okada, Kayoko & Gregory Hickok. 2006. Identification of lexical-phonological networks in the superior temporal sulcus using functional magnetic resonance imaging. Neuroreport 17(12). 1293–1296.

Okada, Kayoko, Feng Rong, Jon Venezia, William Matchin, I. Hui Hsieh, Kourosh Saberi, John T. Serences & Gregory Hickok. 2010. Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. Cerebral Cortex 20(10). 2486–2495.

Oron, A., T. Wolak, T. Zeffiro & E. Szelag. 2016. Cross-modal comparisons of stimulus specificity and commonality in phonological processing. Brain and Language 155–156. 12–23.

Overath, Tobias, Josh H. McDermott, Jean Mary Zarate & David Poeppel. 2015. The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. Nature Neuroscience 18. 903–911.

Peelle, Jonathan E. 2014. Methodological challenges and solutions in auditory functional magnetic resonance imaging. Frontiers in Neuroscience 8. 253.

Peelle, Jonathan E., Ingrid S. Johnsrude & Matthew H. Davis. 2010. Hierarchical processing for speech in human auditory cortex and beyond. Frontiers in Human Neuroscience 4. 51.

Peter, Varghese, Genevieve McArthur & William Forde Thompson. 2010. Effect of deviance direction and calculation method on duration and frequency mismatch negativity (MMN). Neuroscience Letters 482(1). 71–75.

Phillips, Collin, T. Pellathy, Alec Marantz, E. Yellin, K. Wexler, David Poeppel, M. McGinnis & T. Roberts. 2000. Auditory cortex accesses phonological categories: An MEG mismatch study. Journal of Cognitive Neuroscience 12. 1038–105.

Pierrehumbert, Janet. 2002. Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.), Laboratory Phonology VII (Phonology and Phonetics 4.1), 101–140. Berlin: De Gruyter.

Poeppel, David. 1996. A critical review of PET studies of phonological processing. Brain and Language 55(3). 317–351.

Poeppel, David, William J. Idsardi & Virginie van Wassenhove. 2008. Speech perception at the interface of neurobiology and linguistics. Philosophical Transactions of the Royal Society of London B: Biological Sciences 363(1493). 1071–1086.
Politzer-Ahles, Stephen, Kevin Schluter, Kefei Wu & Diogo Almeida. 2016. Asymmetries in the perception of Mandarin tones: Evidence from mismatch negativity. Journal of Experimental Psychology: Human Perception and Performance 42(10). 1547–1570
Praamstra, Peter, Antje S. Meyer & Willem J. M. Levelt. 1994. Neurophysiological manifestations of phonological processing: Latency variation of a negative ERP component time-locked to phonological mismatch. Journal of Cognitive Neuroscience 6(3). 204–219.
Pulvermüller, Friedemann, Yury Shtyrov, Risto J. Illmoniemi & William Marslen-Wilson. 2006. Tracking speech comprehension in space and time. NeuroImage 31. 1297–1305.
Rauschecker, Josef P. 2012. Ventral and dorsal streams in the evolution of speech and language. Frontiers in Evolutionary Neuroscience 4. 7.
Rauschecker, Josef P. & Sophie K. Scott. 2009. Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. Nature Neuroscience 12(6). 718–724.
Rimol, Lars M., Karsten Specht, Susanne Weis, Robert Savoy & Kenneth Hugdahl. 2005. Processing of sub-syllabic speech units in the posterior temporal lobe: an fMRI study. NeuroImage 26(4). 1059–1067.
Rosch, E. 1973. Natural categories. Cognitive Psychology 4(3). 328–350.
Scharinger, Mathias, Alexandra Bendixen, Nelson J. Trujillo-Barreto & Jonas Obleser. 2012. A sparse neural code for some speech sounds but not for others. PLoS ONE 7(7). e40953. DOI: 10.1371/journal.pone.0040953.
Scharinger, Mathias, Ulrike Domahs, Elise Klein & Frank Domahs. 2016. Mental representations of vowel features asymmetrically modulate activity in superior temporal sulcus. Brain and Language 163. 42–49.
Scharinger, Mathias, Carsten Eulitz & Aditi Lahiri. 2010. Mismatch Negativity effects of alternating vowels in morphologically complex word forms. Journal of Neurolinguistics 23(4). 383–399.
Scharinger, Mathias, William J. Idsardi & Samantha Poe. 2011. Neuromagnetic reflections of harmony and constraint violations in Turkish. Journal of Laboratory Phonology 2(1). 99–124.
Scharinger, Mathias & Aditi Lahiri. 2010. Height differences in English dialects: Consequences for processing and representation. Language and Speech 53(2). 245–272.
Scharinger, Mathias, Philip Monahan & William J. Idsardi. 2012. Asymmetries in the processing of vowel height. Journal of Speech, Language, and Hearing Research 55. 903–918.
Scharinger, M., P. J. Monahan & W. J. Idsardi. 2016. Linguistic category structure influences early auditory processing: Converging evidence from mismatch responses and cortical oscillations. NeuroImage 128. 293–301.
Scharinger, Mathias, Jennifer Merickel, Joshua Riley & William J. Idsardi. 2011. Neuromagnetic evidence for a featural distinction of English consonants: Sensor- and source-space data. Brain and Language 116. 71–82.
Schluter, Kevin, Stephen Politzer-Ahles & Diogo Almeida. 2016. No place for /h/: an ERP investigation of English fricative place features. Language, Cognition and Neuroscience 31(6). 728–740.
Schröger, Erich. 2005. The mismatch negativity as a tool to study auditory processing. Acta Acústica 91. 490–501.

Schwarzbauer, Christian, Matt H. Davis, Jennifer M. Rodd & Ingrid Johnsrude. 2006. Interleaved silent steady state (ISSS) imaging: a new sparse imaging method applied to auditory fMRI. NeuroImage 29(3). 774–782.
Schwarzkopf, Dietrich Samuel & Geraint Rees. 2011. Pattern classification using functional magnetic resonance imaging. Wiley Interdisciplinary Reviews: Cognitive Science 2(5). 568–579.
Scott, S. K., C. C. Blank, S. Rosen & R. J. Wise. 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123. 2400–2406.
Scott, Sophie K. & Ingrid S. Johnsrude. 2003. The neuroanatomical and functional organization of speech perception. Trends in Neurosciences 26(2). 100–107.
Scott, Sophie K., Stuart Rosen, C. Philip Beaman, Josh P. Davis & Richard J. S. Wise. 2009. The neural processing of masked speech: evidence for different mechanisms in the left and right temporal lobes. The Journal of the Acoustical Society of America 125(3). 1737–1743.
Scott, Sophie K., Stuart Rosen, Harriet Lang & Richard J. S. Wise. 2006. Neural correlates of intelligibility in speech investigated with noise vocoded speech: A positron emission tomography study. The Journal of the Acoustical Society of America 120(2). 1075–1083.
Seghier, Mohamed L., François Lazeyras, Alan J. Pegna, Jean-Marie Annoni, Ivan Zimine, Eugène Mayer, Christoph M. Michel & Asaid Khateb. 2004. Variability of fMRI activation during a phonological and semantic language task in healthy subjects. Human Brain Mapping 23(3). 140–155.
Sendlmeier, W. F. 1995. Feature, phoneme, syllable or word: how is speech mentally represented? Phonetica 52(3). 131–143.
Shtyrov, Yury, Katja Osswald & Friedemann Pulvermüller. 2008. Memory traces for spoken words in the brain as revealed by the hemodynamic correlate of the mismatch negativity. Cerebral Cortex 18(1). 29–37.
Strand, Fredrik, Hans Forssberg, Torkel Klingberg & Fritjof Norrelgen. 2008. Phonological working memory with auditory presentation of pseudo-words – an event related fMRI Study. Brain Research 1212. 48–54.
Tavano, A. & M. Scharinger. 2015. Prediction in speech and language processing. Cortex 68. 1–7.
Truckenbrodt, Hubert, Johanna Steinberg, Thomas K. Jacobsen & Thomas Jacobsen. 2014. Evidence for the role of German final devoicing in pre-attentive speech processing: a mismatch negativity study. Frontiers in Psychology 5. 1317.
Turkeltaub, Peter E. & H. Branch Coslett. 2010. Localization of sublexical speech perception components. Brain and Language 114(1). 1–15.
Ulbrich, Christiane & Richard Wiese. 2018. Phonotactic principles and exposure in second language processing. In Christiane Ulbrich, Alexander Werth & Richard Wiese (eds.), Empirical approaches to the phonological structure of words (Linguistische Arbeiten 567), 153–182. Berlin & Boston: De Gruyter.
Vaden, Kenneth I., L. Tugan Muftuler & Gregory Hickok. 2010. Phonological repetition-suppression in bilateral superior temporal sulci. NeuroImage 49(1). 1018–1023.
Werth, Alexander, Marie Josephine Rocholl, Karen Henrich, Manuela Lanwermeyer, Hanni Th. Schnell, Ulrike Domahs, Joachim Herrgen & Jürgen Erich Schmidt. 2018. The Interaction of vowel quantity and tonal cues in Cognitive processing: An MMN-study concerning dialectal and standard varieties. In Christiane Ulbrich, Alexander Werth & Richard Wiese (eds.), Empirical approaches to the phonological structure of words (Linguistische Arbeiten 567), 183–211. Berlin & Boston: De Gruyter.

Westerlund, Masha & Liina Pylkkänen. 2014. The role of the left anterior temporal lobe in semantic composition vs. semantic memory. Neuropsychologia 57. 59–70.
Winkler, Istvan. 2007. Interpreting the Mismatch Negativity. Journal of Psychophysiology 21(3–4). 147–163.
Winkler, Istvan , Anne Lehtokoski, Paavo Alku, Martti Vainio, Istvan Czigler, Veleria Csepe, Olli Aaltonen, Ilkka Raimo, Kimmo Alho, Heikki Lang, Antti Iivonen & Risto Näätänen. 1999. Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. Cognitive Brain Research 7(3). 357–369.
Wolmetz, Michael, David Poeppel & Brenda Rapp. 2011. What does the right hemisphere know about phoneme categories? Journal of Cognitive Neuroscience 23(3). 552–569.

Index

!Xóõ 73f., 78f., 88f.

ablaut 103
– ablaut reduplication 129
Abrogans 61f., 66f.
acoustic salience 16, 20
Alemannic 73
analogy 77
aphasia 98
apocope 80
articulatory gesture 4
artificial language 9, 153, 159, 213, 215, 218ff., 222ff., 226ff., 230ff.
aspiration 22, 52
assimilation 13ff., 155, 242

BigBrother-Corpus 20
bisyllable *see* disyllable
bootstrapping account *see* phonological bootstrapping account
boundary strength 16, 18, 28f., 38

CallHome-Corpus 20
Catalan 73, 76, 79, 82
– Central Catalan 6f., 71f., 74f., 79ff., 84, 87ff.
centralization 74, 77, 81, 88f.
Chinese 10, 153, 156, 158ff., 164ff.
– Mandarin 225
– Standard Chinese 129, 158
clicks 78f., 88
Clitic Group 2
clitics 6, 59f., 69
coarticulation 22
Coda Law 73
complex word 5, 13f., 17, 19f., 20, 23, 27f., 37ff., 80, 96f., 99, 102, 114, 116, 254
compound stress 14, 125
consonant epenthesis 77, 80
Contact Law 73
contrast enhancement 184, 190
cortical processing hierarchy 242
cross-linguistic 5f., 68, 71ff., 156, 160, 171, 220, 222f., 229f., 232, 256, 262
cross-modal sentence completion task 104
CV phonology 187

deadjectival nominalization 97, 99f., 115
degemination 3, 13ff.
derivation 7, 13, 89, 96f., 114f., 124
dialect 5, 8, 20, 39, 73, 79, 122, 127f., 183f., 186f., 189ff., 200ff., 258, 261
Dialektintonationskorpus 20
diphthongization 7, 77, 88
disyllable 7, 82, 85ff., 99, 101, 124, 127, 129, 131, 133, 138ff., 205, 219, 221ff., 228
duration 6, 8, 19ff., 23ff., 43f., 74, 162f., 184f., 188, 190, 196, 202, 222, 248
Dutch 37, 82, 219ff., 223ff.
dynamic adaptive system 154

early left-anterior negativity *see* ELAN
EEG 5, 8f., 108, 111, 177, 191, 195, 199, 243ff., 263
ELAN 108, 113
electroencephalography *see* EEG
English 1, 13, 38, 73f., 82, 88f., 97, 103, 126, 137, 155f., 219ff., 223ff., 254
– American English 137, 221
– Canadian English 221f.
– Singapore English 75
episodic representation 241
ERP 8, 97, 102, 108f., 111, 113, 184, 191, 194f., 199ff., 221, 246, 256ff., 262f.
Esperanto 227
Estonian 189
event-related potential *see* ERP
eye-tracking 224, 229

feature 5, 9, 46, 52, 61, 71, 73f., 82, 86, 89, 157, 183, 185ff., 190, 202f., 227, 231, 241f., 252
Fijian 73f.
final devoicing 18f., 258
final lengthening 19, 220
Finnish 89, 189, 216, 220, 223
fMRI 5, 9, 243ff., 254ff., 262f.

foot 2f., 73, 81, 96ff., 125, 128f., 132, 135, 139, 148
foot boundary 148
French 73, 155f., 219ff., 223, 225f.
– Canadian French 222
– European French 222
frequency measure 16, 27
Friulian 88f.
functional magnet resonance imaging *see* fMRI
fundamental frequency 162, 185, 187f.

Georgian 76
German 1ff., 5, 7ff., 13ff., 17ff., 37ff., 45, 61, 63, 67ff., 72, 76ff., 82, 88ff., 95ff., 107ff., 114, 116, 121ff., 132, 136, 141, 153, 156ff., 164ff., 168, 170f., 174ff., 183ff., 221, 223, 225f., 258
– Old High German 6, 45, 47, 51, 53, 61ff., 66ff., 73
– Standard German 8, 183f., 186, 189f., 192ff., 203ff., 211
grammatical word 99, 102, 114
graphematic word 1
graphic demarcation 46
graphic unit 6, 47, 68
graphic word 47
Greek 1, 3

haplology 128
Head Law 73
head-turn preference procedure 228
Heliand 61f., 66
Hiatus Law 73
Hungarian 2, 189

iamb 8, 81, 142, 145, 221f.
Indo-Iranian 73
inflection 7, 23, 89, 96f., 114
information load 19
intensity 162, 188, 193, 222, 231, 233
intonation phrase 19, 24, 26
Italian 77, 130, 221ff.
iterative rhythm 123
Itunyoso Trique 6, 71f., 75f., 79, 83, 85, 87ff.

Japanese 73

Klingon 227
Korean 220, 224
Kulina 76

LAN 109
late positive component *see* LPC
Latin 6, 45f., 51f., 57, 60ff., 67ff., 77
left anterior negativity *see* LAN
lenition 52
lexical access 8, 159, 177, 184ff., 188, 190, 201, 203f., 214, 219
lexical acquisition 214, 217
lexical decision task 155, 229
lexical frequency 17f., 23, 28
lexical stress 81, 116, 131, 147, 216f., 220ff.
lexical tone 188, 190
lexicalisation 64
liaison 225
Liber Evangeliorum 61, 65
loan word 83, 158
Low Franconian
– southern Low Franconian 183
LPC 112f., 115
Luxembourgish 82

magnetoencephalography *see* MEG
mapping rules 2, 71
MEG 243ff.
metrical foot 95, 216
Middle-Franconian 8, 183f.
minimal word 97
minimality constraints 3, 14, 72
Mismatch Negativity *see* MMN
mixed effects linear regression model 24f., 29f., 34, 43f.
MMN 8, 183ff., 191, 195ff., 200ff., 211, 257f., 261f.
monosyllable 7, 64, 97ff., 101, 116, 122f., 125, 129, 131ff., 138ff., 145, 148, 160f., 188ff., 205, 221, 232
mora 2f., 73, 186ff., 191, 201, 204
morphological boundary 15, 21, 28, 37, 125
morphological cues 100f.
morphological word 1
morphophonetics 21
morphophonology 4, 13, 21, 52, 122, 129ff., 141, 146ff., 254

Moselle-Franconian 5, 183f., 186ff., 201f., 204f.
Muspilli 61f., 66

N400 109, 112f., 115, 200, 262
nasalisation 52, 54f.
nonce word 5, 153, 155f., 159ff., 173, 177
non-isomorphism 2, 130
nonword 105f., 223, 226, 229f.
Nucleus Law 73

oddball paradigm 8, 183, 191, 194, 196, 200, 257f., 262
Old Irish 6, 45ff., 51ff., 55ff., 68
Old Saxon 6, 45, 47, 51, 53, 61ff., 66, 68
Onset Maximization 15

P200 109, 112f., 115
P600 113
Palauan 77, 88f.
participle 5, 7f., 95ff., 102ff., 112f., 115
particle verb 1
partitive relation 131, 136f., 147
partword 228, 231ff.
phoneme 3, 6, 65, 69, 86, 157ff., 184, 191, 203, 232, 241ff., 254, 263
phonological asymmetry 242
phonological bootstrapping account 217
phonological feature 5, 9, 241ff., 253, 256, 263
phonological filter 154ff., 159, 165, 176
Phonological Mismatch Negativity *see* PMN
phonological phrase 2, 73, 82, 130, 216
phonological processing 128, 154, 242ff., 247f., 252ff., 262f.
phonological rules 3, 72, 155, 216f., 258
phonological word 1ff., 13ff., 23, 25ff., 36ff., 45, 68, 71f., 74ff., 80ff., 95, 213ff., 218, 227, 232, 234f.
– complex phonological word 23f., 26ff., 36ff.
– simple phonological word 17, 21, 24, 26ff., 31ff., 36f.
phonotactic constraint 2, 5, 155f., 177f., 224, 231, 242
phonotactic restrictions 6, 71f., 75, 81, 88, 223
phonotactics 9, 73f., 85, 153, 155ff., 164f., 176ff., 216f., 224, 226f., 229, 255, 258, 262
phrasal rhythm 123, 130
pitch 9, 113, 186f., 189, 192f., 204f., 220ff., 231, 233
pitch accent 24, 28, 220f.
place name 127f.
plural 52, 59, 80, 97ff., 107, 114ff., 122, 124
– plural formation 96ff., 107, 114ff.
PMN 262
Polish 160, 177
Portuguese 88f.
possessive relation 131, 136f., 147
pre-attentive processing 184, 200ff.
predictive coding 243, 261, 263
Preference Laws 73
prefix addition 107f.
prefix omission 103, 106f.
prefixation 7, 95f., 98, 102ff., 107ff., 113, 115
prosodic constraints 96f., 100ff., 105, 107, 114f.
prosodic cues 8, 99f., 184f., 219ff., 233
prosodic hierarchy 2f., 60, 71, 95
prosodic morphology 96, 121
prosodic parallelism 7, 121, 123, 125, 128f., 131, 138ff., 146, 148
prosodic requirements 102, 106, 109, 114, 116
prosodic template 97f., 103
prosodic word 2, 5, 19, 98, 130, 216
prototype 241
pseudoword 100f., 104, 113, 185, 200, 202, 254
psycholinguistics 7, 95f., 163, 214, 242f.

Quechua 73f.

reaction time 5, 9, 153, 156, 165f., 169ff., 174ff., 219, 229f.
reduplication 129, 148, 227
representational specificity 242
resyllabification 15, 82, 95, 127
rhythmic alternation 7, 121, 123ff., 128, 131, 138ff., 147f.
rhythmic context 131, 135
rhythm-timed language 221

Ripuarian 183
Russian 9, 82, 153, 155f., 158ff., 164ff., 176ff.

Sandawe 78f.
schwa 7, 18, 21, 78, 81, 84, 101, 121ff., 128f., 131, 133, 135f., 141, 145, 147, 220
– schwa optionality 122
– schwa syllable 122, 124, 131, 133, 145
scriptio continua 45f., 62, 68f.
second language acquisition 5, 153f., 159
segment duration 6, 13, 17, 21, 33f., 37f., 192, 231, 233
segmental word-level rules 216
semantic transparency 6, 23ff., 28ff., 32ff., 39
semantic word 1
Smooth Signal Redundancy Hypothesis 16, 39
sonority 9, 153, 156ff., 164, 176ff., 242, 255
sonority hierarchy 156f., 161, 177, 242
Sonority Sequencing Principle 10, 75, 80, 83, 86, 89, 156, 159, 224
Spanish 79, 83
Specific Language Impairment 100ff., 113ff.
speech rate 16, 21, 23, 25f., 28f., 33f., 192
speech register 122
speech segmentation 8, 75, 99, 178, 213ff., 219ff., 229ff., 234, 254
stress assignment 3, 5f., 13, 17, 72, 95
stress clash 7, 121, 123, 126f., 131, 136, 145ff.
stress group 6, 56, 58ff., 68
stress lapse 7, 99, 112, 121, 123, 145, 147
stress pattern 16f., 51, 58, 99, 103, 112, 124, 128, 221
stress-timed language 74, 219, 221
Strict Layer Hypothesis 127, 135f., 148
suffixation 95, 98, 103, 124
Swedish 72, 130
syllabification 15ff., 72
syllable 2ff., 6, 8f., 14f., 19, 21, 23f., 26, 33, 51, 58f., 64f., 71ff., 91, 97ff., 103f., 107, 110, 112f., 115f., 121, 123ff., 131, 133, 137ff., 146ff., 153, 156ff., 160, 177, 186, 205, 216, 219, 221ff., 227, 231f., 241, 252, 254f., 263
syllable boundary 15
syllable language 4, 71ff., 78f., 82
syllable structure 6, 10, 14, 21, 33, 71ff., 79f., 82f., 85f., 88, 131, 141, 232
syllable-timed language 74, 219
syncope 80
syntactic parsing 140
syntactic word 1, 46, 68

Tatian 61f., 66f.
text-to-speech 231
Toda 77
tonal language 84f.
tone 8, 73f., 84f., 184ff., 191, 200ff., 244f., 248
tone accents 8, 183ff., 201, 203ff.
tone-bearing units 187
tone-text-association 8, 184, 186, 188ff., 200ff.
trochee 8, 81, 96ff., 101f., 115, 123ff., 127, 131, 134ff., 138ff., 145, 148, 221f.
Turkish 6, 71ff., 78f., 86ff., 223, 226

underspecification 9, 241ff., 258, 261ff.
universal grammar 154, 156, 178
usage-based approach 154

vowel harmony 2, 7, 9, 78, 86f., 89, 216f., 223, 226, 258
vowel lengthening 77, 84, 88
vowel quality 25, 37f., 124, 189, 191, 202
vowel quantity 183f., 186, 188ff., 194, 196ff., 200ff.

well-formedness 101, 130, 148, 156, 158
Welsh 46, 73
West Greenlandic 73
word boundary 6f., 13ff., 19, 32, 78, 82, 87, 89, 135, 148, 153, 213f., 216, 223f., 227f., 234
word formation 23, 38, 100, 114, 148
word language 71ff., 79, 82, 88f.
– distribution-sensitive word language 6, 72, 85, 88f.
– harmonic word language 6, 72, 87ff.
– stress-sensitive word language 6, 72, 82, 88f.
word separation 45, 50f., 53, 57f., 62, 64, 68
word spotting paradigm 215

word spotting task 219, 225, 229f.
word stress 13, 16ff., 85, 96, 99, 110, 204, 216
word stress rules 216

Yidiɲ 73