Processing German initial consonant clusters

The roles of frequency of use and sonority sequencing in perception and production

Sophia Wulfert

NIHIN studies
New Ideas in Human Interaction

Processing German initial consonant clusters

The roles of frequency of use and sonority sequencing in perception and production

Inaugural-Dissertation
zur
Erlangung der Doktorwürde
der Philologischen Fakultät
der Albert-Ludwigs-Universität
Freiburg i. Br.

vorgelegt von

Sophia Wulfert
aus Berlin

WS 2020 / 2021

Erstgutachter: Prof. Dr. Dr. h. c. Peter Auer
Zweitgutachterin: JunProf. Dr. Adriana Hanulíková
Drittgutachterin: Prof. Dr. Evelyn Ferstl

Vorsitzender des Promotionsausschusses
der Gemeinsamen Kommission
der Philologischen und
der Philosophischen Fakultät: Prof. Dr. Frank Bezner

Datum der Disputation: 06. Dezember 2021

Acknowledgements

Over the course of my PhD, numerous people have contributed to the successful completion of this enterprise. I thank them from the bottom of my heart.

First and foremost, I want to thank my supervisors, Peter Auer and Adriana Hanulíková: for accepting me as their student before they had even met me, for their constructive support in all matters, their patience with me (!), and for always reminding me to focus on the important aspects of the study: your comments have added incredibly much to the structure and content of this dissertation and have made it much more readable. I owe a big thanks to Peter Auer for urging me to be realistic. Without his friendly reminders, I would probably never have finished the experimenting phase. I also wish to thank Adriana for support and mentoring that went far, far beyond the role of a second supervisor and for her unfailing encouragement and positivity: I really appreciate everything you have done for me.

This thesis has been created based on research that I conducted as a member of the Research Training Group (GRK 1624) “Frequency Effects in Language”, and the infrastructure provided by the project has been extremely helpful. More importantly, however, I would like to thank our coordinators Michael Schäfer, Nikolay Hakimov, and Agnes Schneider-Musah, the professors associated with the group, and my colleagues at the GRK. It has been both inspiring and very, very enjoyable to work with you, and am immensely grateful for having been a member of this amazing group. Some of my colleagues I would like to give a special mention: I thank Udo Rohe for being a wonderful colleague and friend: the lunches with you (especially in the final phase of the GRK) were a daily highlight for me. The discussions about linguistic and non-linguistic issues were as delightful as thought-provoking. Luke Bradley and Emiel van den Hoven made numerous

insightful suggestions on my project and were a constant source of knowledge. Thank you for your support and the willingness to share your wisdom. I had the pleasure of sharing my office with Annette Fahrner, Karolina Rudnicka, Helena Levy, and Boniface Nkombong, who were great company and made the work much more enjoyable. Annette, thank you for being a wonderful office mate, friend, and role model (your calm determination and diligence always set a good example) and for cheering me on towards the end. Laura Terassa, Ana Estrada, Uliana Schöller, Miriam Burk, Olga Glanz, Pia Hagen-Wiest, Bella Diekmann, and Laura Cuthbertson were loyal colleagues and friends, who were always ready to help and fun to be around.

But most of all, my heartfelt thanks go to Carmen Pietropaolo and Helena Levy, the best friends and companions I could ever have wished for! Thank you for going through this life phase with me, for precious advice on various matters, and for your moral support in times when I most needed it. Words do not express how grateful I am for having had you by my side and for sharing joy and difficulties with you. Helena, it has been a pleasure to share so much of this phase with you, from participation at the winter school to workshop organisation and teaching our first class of undergraduates.

At various stages of this project, people contributed to the planning, execution, and analysis of my research, as well as the actual writing phase. Laura Terassa, Udo Rohe, Annette Fahrner, Helena Levy, Carmen Pietropaolo, Miriam Burk, Georg Klugmann, Patricia Huber, and Monika Nitz volunteered to take pretests of my experiments and gave me valuable feedback on them. The Medienzentrum at the University of Freiburg made their audio booth available to me, and the staff were always helpful and cheerful. Taylor Veinotte and Katharina Balthasar assisted me in recruiting experiment participants and in data annotation. Göran Köber, Anne Krause, Lars Konieczny, Christoph Wolk, and Carmen Pietropaolo taught me everything I know about statistics and helped me with various aspects in the interpretation of my data. Marten Juskan taught me the magic of LaTeX and readily answered all my questions, even years after the official course. My writing buddy

Gert Herold made the last writing phase so much more enjoyable and was a great partner to discuss all practical matters.

I am much indebted to Laura Cuthbertson, Mirka Honkanen, Helena Levy, Carmen Pietropaolo, Sarah Warchhold, Anna Barkhoff, and Annette Fahrner, who kindly proofread (long!!) parts of this dissertation (or the whole dissertation, in Laura's case) and provided very valuable and much appreciated comments on how to improve it on so many levels. A big, big thank you to all of you! Furthermore, the project certainly benefited from the discussions at colloquia and helpful comments on my presentations.

Without financial support, this research project could not have been carried out. I wish to thank the Deutsche Forschungsgemeinschaft (DFG) for the funding I received as a member of the Research Training Group and to thank the Internationale Graduiertenakademie (IGA) for a scholarship that allowed me to complete the dissertation project. I am also immensely grateful for the opportunity to visit the MARCS Institute at Western Sydney University and to profit enormously from the meetings there. In particular, I would like to thank Anne Cutler, who sadly was taken from us before this dissertation got published, for welcoming and supervising me in Sydney and for very inspiring and fruitful discussions (as well as ice cream with a magnificent view from her balcony). Furthermore, my thanks also go to Laurence Bruggeman, Varghese Peter, and the researchers at Macquarie University, as well as the tech team at Western Sydney University for their support. Without you, the L2 experiment would not have been possible.

Finally, a number of people have accompanied me on my professional and personal journey. Karl Maroldt gave me the most fascinating and engaging start into the world of linguistics imaginable and always supported me in pursuing my academic goals. Nikolay Hakimov has been an inspiration and very supportive both professionally and emotionally. I also want to thank Georg Klugmann for his constant willingness to help with everything from technical problems to theoretical considerations and for his brisk outside viewpoint, but most of all for his moral support through most of this period and for always believing in me, even when I did not believe in myself. Moreover, I

am grateful to Yoav Nahshon for pythonising tasks where my R skills failed me (as well as helping with other technical issues), for pushing me on during the last phase of my PhD, and for showing me—and constantly reminding me—that there is a life beyond the thesis and a world outside my window.

Last but certainly not least, my cordial thanks go to my parents, who have supported me unquestioningly in absolutely everything I have done, who have always had an open ear for all my ups and downs, and who have shown a keen interest in my progress and everything related to the PhD project (even though they will probably never fully understand my fascination with linguistics). I dedicate this dissertation to them.

Contents

List of Figures

List of Tables

1. Introduction

The human speech processing system is highly efficient and automatised. Although speech production and perception are extremely complex (for example, speech articulation might be the most complex behaviour humans perform, cf. Lieberman, 1985) and still not completely understood, every healthy human can communicate effortlessly in his or her native language and does so, most of the time, without consciously thinking about it. In everyday communication, speakers produce an average of four to six syllables per second (Reetz & Jongman, 2009)—and, consequently, listeners perceive and interpret the same number of syllables. Errors are rare, amounting to less than five per 1000 spoken words (Garnham et al., 1982; Leuninger, 1993).

To achieve this degree of efficiency, the human speech processor makes use of whatever information it has access to, including structural regularities in language, in order to guide processing. The automatisation of speech perception processes by means of available cues saves cognitive resources (Segalowitz & Segalowitz, 1993). For example, expectations concerning upcoming elements, based on syntactic frames and/or semantic context, are utilised for rapid speech perception and disambiguation (e.g., Borsky et al., 1998; Marslen-Wilson, 1975). Furthermore, rhythmic regularities have been found to speed up speech processing (for perception: Cason et al., 2015; for production: Tilsen, 2011).

On a sublexical level, regularities[1] in sound sequencing can be used in a similar fashion. For example, listeners use restrictions on se-

[1] See below for an explication of what *regularity* entails.

quences like the Obligatory Contour Principle (Frisch, 2004), vowel harmony (Kabak et al., 2010) and preferences for consonant sequences (Zhao & Berent, 2016), and phonological alternations (Warner et al., 2005) to guide processing; likewise, such regularities are also exploited to facilitate speech production (Carré et al., 1995; Cholin & Levelt, 2009). In comprehension[2], the ability to use expectancy and top-down information to guide speech processing is useful for repairing imperfect input, which arises from production errors, a noisy channel, or perception errors. In such cases, making use of structural knowledge helps in determining the most likely interpretation of the speech input. This ability can lead to perceptual illusions, for instance, in the form of epenthetic vowels between consonants which the listener knows to constitute an illegal cluster (e.g., Wagner et al., 2012). The fact that phonotactic violations are detected automatically prior to conscious recognition (Steinberg et al., 2011) suggests that perceptual repair is not the result of a deliberate repair strategy but is similarly automatic in nature. From the above, it follows that less regular patterns or patterns that are less "good" should be harder to process, leading to longer processing latencies and lower accuracy. Furthermore, knowledge of sublexical regularities can also be utilised to make processing of correct input more efficient.

It is of theoretical importance to identify the factors utilised by language users to guide speech production and perception. On the lexical level, word frequency (Oldfield & Wingfield, 1965), familiarity (Gernsbacher, 1984), imageability/concreteness (Cortese & Schock, 2013; although see Connell & Lynott, 2012), and more have been identified to facilitate processing. On the sublexical level, the phonological composition of words influences their processing. The question arises as to which properties in particular benefit processing. Both universal and

[2] Since this dissertation is concerned with spoken language, the terms *perception* and *comprehension* will be used to refer to auditory perception and the processing steps that follow, respectively, throughout the whole thesis.

language-specific factors have been found to influence processing. For instance, words consisting of universally preferred syllable types, like CV, are easier to process than words consisting of dispreferred syllable types, such as CCV (Brendel et al., 2008). Moreover, learnability and processing are facilitated for syllables that conform to sonority sequencing (Ulbrich et al., 2016), that is, syllables that display a rise in loudness/vocal opening towards the nucleus and a decline thereafter. Sonority sequencing is argued to be a universal concept (Selkirk, 1984; see also Section 3.3). Aphasic speech errors, as well as aphasic neologisms, also indicate a preference for syllables that conform to sonority sequencing and report processing difficulties for syllables that violate it (Miozzo & Buchwald, 2013; Stenneken et al., 2005).

In addition to these universal factors, language-specific regularities affect speech processing. Evidence suggests that phonotactic knowledge, that is, knowledge of which segments can occur in which syllable position and in combination with which segments according to the rules of a specific language, exerts a strong influence during automatic speech processing. Phonotactic restrictions of a language user's mother tongue impede both perception and production of illegal phoneme sequences. In production, for example, non-native consonant clusters are often produced with an epenthetic vowel (Tajima et al., 2003; Yazawa et al., 2015). Likewise, such sequences are also difficult to perceive, which leads to the perception of an illusory epenthetic vowel (Dupoux et al., 1999). Not only is the processing of phonotactically impossible sequences inhibited, experience with native-language phonotactics also affects processing in a gradient manner (probabilistic phonotactics)[3]: processing of common sequences (e.g., syllables or biphones) has been found to be facilitated. This means they are produced and perceived faster and/or more accurately compared to less

[3] The statistical distributions of phonotactic patterns will be referred to as *gradient phonotactics* throughout this thesis, which stands in contrast to *categorical phonotactics*, the legality status of phonotactic patterns.

frequent sequences (Edwards et al., 2004; Levelt & Wheeldon, 1994; Vitevitch & Luce, 1999). There is evidence that listeners are biased towards perceiving the more probable segment sequence—given a specific context—in ambiguous listening situations (Pitt & McQueen, 1998). However, both universal and language-specific effects of phoneme sequencing depend on the specific processing setting (task requirements, population, and stimulus material), the details of which are not determined conclusively. Moreover, most studies examine only one such effect so that any interplay cannot be observed.

It is the aim of this dissertation to shed further light on the relative contributions of universal and language-specific sequencing preferences (represented by sonority sequencing and consonant cluster frequencies, respectively) to the facilitation and automatisation of speech perception and production, specifically that of consonant clusters. This pertains to the perpetual debate of nature vs. nurture. On the one hand, a connection is often drawn between the cross-linguistic distribution of phonotactic sequences and their ease in speech processing (e.g., Goldrick, 2002). This connection follows from the logical assumption that what is easy to process will be more likely to survive in the world's languages; this parallels Jakobson's (1962) original claim that the cross-linguistic distribution of sounds, the order of acquisition in child language, and the order of loss in aphasia, as well as susceptibility to speech errors, are related ("nature"). On the other hand, it is plausible that such natural biases can be overcome and replaced by learned preferences, so that a cross-linguistically dispreferred phoneme sequence may be preferred by speakers of a language in which it is very common ("nurture"). This is due to the over-learning of the sequence by speakers of that language. This rationale is anything but new. Already Aristotle noted in his *De Memoria et Reminiscentia*: "[H]abit [...] takes the role of nature. [...] and frequency makes it nature." (Aristotle, ed. by Bloch, 2007; 452a).

The question is: to what extent does nature guide processing and at what point does habituation take over? Infants, for example, are able to discriminate between all sound distinctions at birth but become insensitive to contrasts that are not phonemic in their native language (L1) at around six months of age (Best, 1994; Kuhl et al., 1992). They then attune to their native language's phonotactics about two months later (Aslin et al., 1998).

It is also of theoretical interest what happens when another language comes into play. In second language (L2) processing, there are three potential sources for processing biases: universal preferences, L1 phonotactics, and L2 phonotactics, each of which might contribute to processing to different degrees. Naturally, in many domains of L2 processing, universal principles and the structure of the L1 have the greatest influence at the beginning of L2 acquisition; with growing proficiency in the target language, its structure becomes more influential in guiding processing (for syntactic processing, cf. Lenzing, 2015; Seibert Hanson & Carlson, 2014; for decontextualised word reading, cf. Chikamatsu, 2006; Miller, 2011). As regards the use of L2 phonotactic knowledge, advanced learners are able to use it to speed up word recognition; conversely, beginner learners only show effects of L2 phonotactics when their L1 allows a greater variety of phonotactic patterns than the target language (Trapman & Kager, 2009). This can be taken as an indication that the positioning of the L1 and L2 with regard to universal phonotactic preferences influences L2 processing at low proficiency levels but that more experience with the L2 (and thus higher familiarity with L2 phonotactic patterns) can nullify this effect. The exact interplay of universal principles, like sonority sequencing, and language-specific phonotactics in L1 and L2 speech processing is nevertheless still unclear.

The present dissertation investigates this issue on the basis of German consonant clusters of varying frequencies and varying degrees of conformity to sonority sequencing. The same set of 16 consonant clusters (/ts/, /ʃt/, /ʃp/, /tr/, /kr/, /ʃl/, /fl/, /ʃm/, /pl/, /ʃn/, /sk/, /ps/, /sl/, /tʃ/, /ks/, and /sp/) is utilised in native and non-native perception experiments, as well as a native production experiment, in order to directly compare effects across tasks and populations, and to draw conclusions on how these factors modulate effects of frequency and sonority in sublexical processing. The same factors are not necessarily influential in production and perception (let alone in L1 and L2 processing)—as Kabak and Idsardi (2007) noted, "perceptual phenomena are not simple inversions of phonological phenomena in the production system". It is therefore insightful to compare effects in the two modalities, and in L1 and L2 processing. Initial consonant clusters are used because the speaker–hearer is most susceptible to general phonotactic influences in such cases and cannot be influenced by any phonemes prior to the consonant cluster.

It is generally difficult to de-correlate frequency and sonority in phonotactic structures because phonologically marked structures, such as sequences that violate sonority sequencing, tend to have a low frequency in most languages (Frisch, 2015). However, initial sibilant–stop sequences are a well-known exception, since they violate sonority constraints but are relatively common in a number of languages, including German.[4] They are therefore valuable test cases in which sonority-based markedness and frequency diverge, while simultaneously serving as an opportunity to de-correlate frequency and sonority sequencing.

Since frequency effects in language acquisition and processing, as well as many other domains of cognition in humans (and animals), have been widely acknowledged and are hardly ever disputed, it might be asked what this study contributes to the understanding of the mechanisms of speech processing: in some sense, it could be seen as stating the obvious. For example, Yang (2015) notes that, while frequency

[4] For a discussion of the applicability of sonority constraints to these clusters, see Section 3.3.3.

effects are ubiquitous, they are of little theoretical importance; he instead argues for structural effects, for example on language change. The goal of the present study is not so much to argue for human sensitivity to frequencies in general—its existence has been sufficiently established—but rather to explore the extent of that sensitivity: to what kinds of linguistic units does it apply and how strong is the effect? The exact configurations under which it pertains and its interaction with other factors, mainly sonority sequencing, will be investigated. In this regard, the study at hand is of definite theoretical importance.

The main research question this dissertation seeks to answer is to what extent the frequencies of consonant clusters and their adherence to sonority sequencing principles influence their processing, respectively. Are cluster frequencies more influential in facilitating speech processing than the universal structuring principle of sonority sequencing and (how) do the two interact? More specifically, the following questions are addressed:

1. Do the frequencies of initial *consonant clusters* influence speech processing (production and perception) over and beyond the frequencies of their subordinate (the clusters' component *phones*) and superordinate units (the *syllables* whose onsets the clusters constitute)? This relates to whether consonant clusters are relevant units in speech processing.

2. Does sonority sequencing in consonant clusters influence their processing?

3. Do the (relative) contributions of these two factors differ between speech production and speech comprehension?

4. Do L1 and L2 listeners exhibit different patterns of sensitivity to sonority sequencing vs. target-language-specific phonotactics? How strong is the influence of L1 phonotactics compared to that

of L2 (i.e., target language) phonotactics and sonority sequencing?

In this dissertation, three experimental studies are reported in which processing difficulty is operationalised as production and perception errors. They will therefore also reveal whether speakers and listeners are sensitive enough to frequency and sonority sequencing for facilitation to surface as diverging accuracy rates in the recognition of different consonant clusters.

The thesis is structured as follows: The next two chapters cover the theoretical background of the thesis by giving introductions to both usage-based linguistics and consonant clusters as units in phonological theory and psycholinguistics, with a special focus on the concept of sonority sequencing. Chapter 3 furthermore presents the specific consonant clusters used throughout the experiments reported in this dissertation. Chapter 4 summarises the most important prelexical steps and phenomena in speech perception, in particular the perception of consonants and consonant clusters, and introduces two connectionist speech perception models. In Chapter 5, an L1 identification-in-noise experiment is presented, followed by a discussion of the results. Chapter 6 reports an experiment in which the same method is applied to a group of L2 listeners; the results are then discussed and compared to those of the L1 listeners. Chapter 7 provides an overview of speech production processes and reports the most important findings of previous studies. It also expounds on the utility of speech errors for research on speech production processes and mechanisms, and presents one of the most influential connectionist models of speech production. In Chapter 8, a speech production experiment involving a tongue twister task is presented and the implications of its results for speech production are discussed. The results of all three experiments are compared in Chapter 9, and their implications for the processing of sublexical units and consonant clusters, in particular, is discussed. Finally, the main results and conclusions from all experiments are summarised in Chapter 10.

2. Usage-based linguistics and frequency of use

The theoretical background to the frequency-related hypotheses of this study lies in usage-based linguistics. In the current chapter, this theoretical framework and its most important premises are laid out.

2.1. Usage-based linguistics

Usage-based linguistics is rooted in cognitive linguistics and departs from traditional conceptions of language, held by generativists, according to which linguistic competence and performance are inherently different. In contrast, grammar and language use mutually shape each other from a usage-based point of view. They are indivisible. Grammatical knowledge is based on knowledge of language usage and generalisations over several usage events (Ibbotson, 2013). Acquiring it involves a process of statistical inference; language learners "keep track of co-occurrences among features of linguistic stimuli" and learn their predictive dependencies (Kapatsinski, 2014; p. 5). In contrast to generativism, usage-based linguistics assumes that language users need very little a priori knowledge in order to accomplish this kind of acquisition. Instead, it is general cognitive processes, which are also used in non-linguistic tasks (and by animals), that are deployed both in language learning and use. It is the aim of usage-based theories of language to explain language use and linguistic structure in terms of domain-general cognitive processes. Bybee (2010) lists five such processes that characterise linguistic processing and, ultimately, create linguistic structure: 1) categorisation, 2) chunking, 3) rich memory, 4) analogy, and 5) cross-modal association.

Categorisation, the grouping of similar entities into the same bin, is used in a large number of cognitive domains, such as recognising the visual input of a *Lanius senator* outside the window as a *bird* due to its similarity with the many other birds one has previously seen. In speech perception, it refers to a matching process between words or phrases in the input and stored representations in memory. If somebody were to point to the entity in this example and say, "Look at the bird!", the speech signal [bɜːd] would be matched with a representation that contains, inter alia, semantic concepts associated with birds more generally. Categorisation is considered the most pervasive of the five processes because it interacts with all of the other types (Bybee, 2010).

Chunking is a process whereby sequences of units that frequently occur together fuse to form a larger unit, a chunk that is stored and processed holistically (see below for examples). Outside of speech processing, it is used in the execution of movement sequences (e.g., Verwey & Abrahamse, 2012), for example. Chunking can be either deliberate (for instance, as a mnemonic strategy in recall tasks) or automatic; automatic chunking is the most relevant type in usage-based linguistics.

Rich memory refers to the level of detail with which events and impressions are stored.

Analogy, a relational matching between entities, is used to produce novel utterances based on familiar ones, for example, a sentence with the same syntactic structure as one heard earlier but consisting of different lexical items. Analogy is a central component of human cognition and is used, for instance, in the processing of geometrical shapes (Holyoak et al., 2001).

Cross-modal association refers to the ability to link form and meaning. In general, it is responsible for the joint storage of co-occurring experiences.

According to Bybee, linguistic structure emerges from the repeated application of these five processes (i.e., the conventionalisation of patterns of usage). Since it is repetition that shapes representations, frequencies of use of linguistic units are of central importance in usage-based linguistics and are thought to have a huge impact on both the processing of these items and the formation of new patterns in language change. Crucially, it is assumed that each experience with language, and every encounter with a specific token (e.g., a phrase, a word, or a phoneme) has an impact on its representation in memory. The encounters are stored as so-called *exemplars*, which are arranged in *exemplar clouds* of similar tokens. The exemplars are stored with great detail, including fine phonetic detail, contextual information, and inferences. This is what *rich memory* (point 3 above) refers to. Categorisation is used to map these rich memories onto representations. Each encounter with a token has an impact on its representation because it either strengthens an existing exemplar or adds another one to the cloud (Bybee, 2010). Via the strengthening of existing representations, frequency of use facilitates their activation and processing (Diessel, 2017).

Frequency also has an impact on processing by means of *chunking* (point 2 above). As previously mentioned, frequent co-occurrence of items in a sequence subjects them to chunking. Bybee (2010) points out that meaning is then assigned to the largest chunk that is available to the human processor. However, she stresses that, “even though chunks are stored as units, their constituent words are still closely related to the general exemplar clusters for those words.” (Bybee, 2010; p. 42) As is apparent from the above quote, the focus of chunking in the majority of the usage-based literature is on several words becoming one unit, such as the frequently uttered *I don’t know* becoming one chunk that is stored and processed as a whole, which results in the phonetically reduced *I dunno* or similar forms. This dissertation will address whether and to what degree this process also applies to phonemes' forming larger chunks, namely consonant clusters (see also Section 3.5.1). Of the processes mentioned above, chunking is therefore the most relevant process to the present study.

Closely related to chunking is the psychological process of entrenchment, which represents one of the central concepts in usage-based linguistics.[1] *Entrenchment* refers to the process of a highly complex event becoming a "well-rehearsed routine that is easily elicited and reliably executed. When a complex structure comes to be manipulable as a 'pre-packaged' assembly, no longer requiring conscious attention to its parts or their arrangement ...it has the status of a unit"(Langacker, 2000; pp. 3–4). Entrenchment can operate on several linguistic levels, which means that it involves structures of different sizes, such as phoneme or word sequences. For example, concerning the composition of the mental lexicon, Bybee (1999; p. 232) writes: "[T]here is a set of highly entrenched gestures and gestural configurations that are used and re-used in constructing the words of a language." As can be inferred from this wording ("highly entrenched"), entrenchment is gradual: units are variably entrenched as a function of their frequency of occurrence (Langacker, 1987). Structures that are processed as a holistic unit are at one end of this spectrum. In this sense, chunks can be said to represent extreme cases of entrenchment. From a neuro-cognitive point of view, entrenchment can be accounted for in terms of the *Hebbian learning rule* (Hebb, 1949): "What fires together, wires together", which means that the connections between concurrently active nodes are strengthened. In addition to frequency, age of acquisition has also been found to be a relevant factor for entrenchment (Baumann & Ritt, 2018). For an in-depth discussion of entrenchment in usage-based the-

[1] In fact, Langacker (2000) includes entrenchment as one of five domain-general cognitive processes relevant to language in a listing similar to that of Bybee's above.

ories, see Blumenthal-Dramé (2012). Acquiring form–meaning pairings is central to any linguistic learning process. In usage-based approaches (especially in Construction Grammar, CxG, cf. Goldberg, 1995), the acquisition of grammar consists in making a connection between a generalised form, which is derived from the many instances stored from previous experience, and its meaning (*cross-modal association*, point 5 above).

As Kapatsinski (2014) points out, however, this form–meaning association does not exclude the possibility of language-acquiring children (and also adult speakers) tracking and learning purely form-based relations with no direct connection to meaning. Speakers can extract phonotactic patterns to predict upcoming units, for example. Grammar on all levels is supposed to be stochastic in nature, so to learn a grammar is basically probability matching (i.e., keeping track of distributions and applying this knowledge to production and recognition), much like in non-linguistic learning processes (Kapatsinski, 2014). Here, it one again becomes apparent that the usage-based approach aims to situate linguistic behaviour within human cognition in general and to explain it in terms of the same mechanisms as behaviour in other cognitive domains.

From the above, it follows that grammar according to usage-based linguistics is not autonomous (as postulated by nativism) but a dynamic, adaptive system (Bybee, 2010) "that emerges from frequently occurring patterns in language use" (Diessel, 2003; p. 167). It is conceived of as "the cognitive organisation of one's experience with language" (Bybee, 2010; p. 8). Hence usage, cognition, and the structure of a language interact, according to the usage-based approach.

In the next section, frequency of usage will be defined, in addition to an outline of the importance of frequency of use of linguistic units and statistical learning within the framework of usage-based linguistics. Furthermore, empirical support for this stance will be provided.

2.2. Frequency of use and its effects in psycholinguistics and language change

2.2.1. Frequency measures and their definitions

Frequency of use, straightforward as it may seem, is not a uniform concept but can rather refer to a number of distinct measures. Spoken frequencies may diverge from written ones, and language users' subjective frequencies seem to be separate for the two modalities (Gaygen & Luce, 1998). Subjective frequency is also related to the measure of familiarity. In fact, frequency and familiarity are often correlated (Gernsbacher, 1984).

Most importantly, however, type and token frequencies have to be distinguished. While in some cases, the effects of the two kinds of frequency are similar, in others they can be antagonistic. *Token frequency* refers to the number of occurrences of a linguistic item, which is usually derived from text corpora. For example, the token frequency of the German word *Abenteuer* according to the CELEX database[2] is 132, meaning the word occurs in the corpus a total of 132 times. Often, token frequencies are given per one million words. As this version of the Mannheim corpus consists of 6.0 million words, *Abenteuer* has a frequency of 22 per one million words. *Type frequency* on the other hand, denotes the number of different lexemes that are associated with the entity under investigation. On the level of the construction[3], the most common application of type frequencies in usage-based theories, type frequency identifies the "number of distinct lexical items that can be substituted in a given slot in a construction" (Ellis and Collins, 2009; p. 330). An example is the number of different verbs that can take the regular past tense morpheme *–ed* in English. On a sublexical level,

[2] based on lemmas in the Mannheim corpus, spoken and written frequencies (Baayen et al., 1995)

[3] a form-meaning pairing containing at least several morphemes

however, type frequency refers to the number of different lexemes (i.e., higher-level units) a unit *occurs in*, which can be determined from a dictionary. For example, the CELEX type frequency of the German syllable /vɪl/ is 18, as it is a constituent of 18 different lemmas (e.g., *willkommen*, *willkürlich*, *Pavillon*, *verwildern*, etc.). Its token frequency, which is calculated by summing the individual token frequencies of the 18 lemmas, is 381. Type and token frequencies of sublexical units are not necessarily correlated, although it has been found that high-frequency (HF) words tend to be composed of more common biphones than low-frequency (LF) words (Frauenfelder et al., 1993). For the consonant clusters used in the experiments here, for example /tr/ and /kr/ have relatively similar token frequencies, whereas the type frequency of /tr/ is almost twice as high as that of /kr/. Similarly, /ʃn/ has a higher token frequency than /ʃm/ but the ranking is reversed when it comes to type frequencies. (See Section 3.5.1 for plots of type and token frequencies of all test clusters.)

An important question to ask is what kinds of frequencies our phonotactic knowledge is built upon, mainly whether type or token frequencies are the more relevant influencing factor. There is theoretical rationale for both kinds of measures. Since type frequencies reflect the number of different lexemes associated with a construction or another entity, they are generally related to pattern strength and productivity (Albright, 2009; Bybee, 2010). This is because its abstraction is promoted by the larger number of examples that feature the item—the generalisation itself is strengthened rather than the individual examples. This has been discussed in detail with regard to constructions (e.g., Bybee, 2010). Constructions with a high type frequency are not as closely associated with any particular lexeme and are therefore more easily generalisable to other lexemes (Archer & Curtin, 2011); they have a high productivity. This effect of generalisability is specific to type frequency.

High token frequency, on the other hand, fits better into the exemplarist picture of a network shaped by the summed encounters with a particular structure in language. High token frequency leads to the strengthening or entrenchment of a particular instantiation of a linguistic unit. It therefore increases the autonomy of that unit, which probably develops its own representation (Archer & Curtin, 2011) and is increasingly accessed directly in speech production and perception. As result, items with a high token frequency contribute less to the productivity of the constructions (or words) in which they are contained. Due to its autonomy, token frequency is often associated with lexical access and online processing, while type frequency is the more relevant measure for grammar. However, as Albright (2009) points out, the distinction is by no means clear: not all online tasks are sensitive to token frequencies and some grammatical tasks are. For the tasks related to sublexical processing employed in the present studies, it seems plausible that type frequencies are more relevant because nonce words are used as stimuli in all experiments. Therefore, there are no individual lexical representations to draw on but rather pattern frequencies, which are best represented by type frequencies. Hay et al. (2004; p. 61) also come to the conclusion that it is type frequencies and not token frequencies that are "most directly related to phonotactic well-formedness".

Both type and token frequency have been tested for effects in speech perception, with both showing effects in a wide range of tasks and conditions (type: e.g., Cheng et al., 2014; Luce & Large, 2001; token: e.g., Jusczyk et al., 1994; Vitevitch, 2003). As the two frequency measures are often strongly correlated[4] (and tested separately), it is difficult to discriminate between their relative influence. There are converging indications, however, that type frequencies are better predictors of

[4] As Berg (2014; p. 199) demonstrated, this correlation is restricted to phonology, while "[n]on-phonological distinctions evince a higher discrepancy between type and token frequency [...]."

processing performance than token frequencies. Some studies specifically designed to test token frequencies have failed to find effects (e.g., Warner et al., 2005; in perception of gated biphones; Archer & Curtin, 2011; in infant speech processing). In the few studies that directly compared the effects of type and token frequencies, type frequencies outperform token frequencies, although both influence processing in some cases (Hay et al., 2004; Hayes & Wilson, 2008; Janse & Newman, 2012; without token effect: Archer & Curtin, 2011). It is therefore conceivable that both frequency measures influence speech processing, but that type-based phonotactic probability exerts a stronger effect than its token-based counterpart. Note, however, that in a model of generalised phonotactic acceptability ratings by Albright (2009), taking token frequencies into account even resulted in deteriorated model performance.

Apart from type and token frequencies, transitional probabilities (i.e., the probability of one phoneme, syllable, or word given an adjacent one) have been used as a measure in psycholinguistic experiments and have proven to show effects on acquisition and processing (e.g., Aslin et al., 1998; Yip, 2000). In addition to forward transitional probabilities, listeners as well as language learners make use of backward transitional probabilities. In fact, it has been shown that backward transitional probabilities make *better* predictions concerning word-likeness judgements than forward transitional probabilities (Perruchet & Peereman, 2004).

With respect to consonant clusters, there has been considerable debate concerning the question of whether the specific frequencies of a particular sequence or generalised frequencies are more influential in certain tasks (e.g., Albright, 2012; for morphological productivity). If sequence-specific frequencies are the most relevant measure, then the German initial cluster /tr/ should be processed very easily since it has a high phonotactic probability as a German consonant cluster, while /sl/ should be more difficult to process since it has a low phonotactic probability. The impact of language-specific gradient phonotactics might not be as straightforward as that, however. Particularly behavioural differences with regard to unattested consonant clusters (i.e., with equal, namely zero, probability) pose problems for such an explanation. Several researchers (e.g., Albright, 2009; Daland et al., 2011; Linzen & Gallagher, 2014) argue in favour of a generalisation process in which the output, generalised probabilities, is the source of preferences for certain phoneme sequences. According to them, concrete phoneme distributions are abstracted over, most probably on the basis of features or natural classes, to arrive at more general phonotactic evaluations. The advantage of these generalised probabilities is that they apply to attested and unattested phoneme sequences alike, although they usually only outperform segment-based models with respect to unattested clusters.

2.2.2. Frequency effects in linguistics and psycholinguistics

Frequency effects are observable both in the psycholinguistic domain (in L1 as well as L2 acquisition, language production, and language comprehension) and in the domain of language change. As Bybee elucidated (see above), this is because the two are so intimately connected—the latter can be interpreted as a consequence of the former. In speech processing, frequency effects have been reported since the advent of psycholinguistic research, and several publications have been devoted specifically to the effects of frequency in different linguistic domains (e.g., Ambridge et al., 2015; Behrens & Pfänder, 2016; Diessel, 2007; Ellis, 2002). Frequency effects have been shown to involve units of different sizes, such as words and phrases, but also sublexical units like phonemes and syllables. This section gives an overview of the most prominent types of frequency effects and their mechanisms.

Basically, repeated use of a linguistic item leads to increased learnability, facilitation in processing, and—with a restriction to be discussed later—diachronic preservation. In language acquisition, repetition (i.e., high input frequency) is one of the main determinants of successful learning. It is through repeated encounter that children acquire both concrete lexemes and more abstract constructions, such as syntactic and phonotactic patterns (which are generalisations over a number of concrete instantiations). That is why, all other things being equal, frequent items are acquired earlier than infrequent items (Moerk, 1980). However, the relationship is intricate and not straightforward (Goodman et al., 2008). For example, it is not mere token frequency but the interaction between type frequency and skewed type–token ratios in the input that determines the learnability of an L2 construction (Madlener, 2016).

Since usage-based linguists view linguistic knowledge as a dynamic system that is constantly changing and evolving, they hold that input frequencies continue to influence mental representations and their activation. Both in speech perception and production, there is strong evidence for facilitated access to HF items and sequences in the mental lexicon. In perception, this shows as faster and more accurate recognition of linguistic items with a high frequency of occurrence (Grosjean, 1980). For example, in noisy listening situations, HF words are recognised more reliably than LF words (Howes, 1957). Similarly, in comprehension errors, the percepts tend to be of a higher lexical frequency than the targets (Savin, 1963). In production, frequency effects are evident in shorter latencies for the preparation of HF items (Oldfield & Wingfield, 1965) but also in multi-word sequences: HF sequences are interrupted by hesitations less often than LF sequences (Schneider, 2014), which shows reduced planning effort for these items as a chunk. All of this suggests facilitated access to HF words and longer constructions.

Another effect in production is the phonemic and phonetic reduction of HF items, which includes the phonetic shortening and deletion of single phonemes in lexemes (Gregory et al., 1999; Jurafsky et al., 2000), as well as the contraction of longer HF chunks, such as *I dunno*. These reduction effects demonstrate that the reduced effort associated with the production of HF elements is not limited to the planning stage but extends to articulation, too. Such effects can be explained by a simple practice effect in production (Diessel, 2007) but probably also bear witness to the interplay between the two antagonistic forces in speech production: articulatory economy and the necessity for intelligibility. HF words and phrases are more likely to be reduced because they are more predictable and their recognition does not depend on careful pronunciation as much as it does for LF words and phrases; this allows the speaker to minimise articulatory effort while at the same time retaining intelligibility. This predictability-based interpretation is supported by the fact that this applies to lexeme frequency as well as conditional probabilities between words (Jurafsky et al., 2000). Moreover, frequent words are reduced more in contexts in which they most frequently occur; according to Bybee (1999), this is due to the fact that they are part of a larger processing unit in such contexts.

The reduction effect is a particularly good example of how effects arising in the cognitive and articulatory needs of the communication situation become conventionalised over time and lead to changes in the language system: many reduced forms (such as *I dunno* mentioned above or *gonna*, but also reduced single lexemes like *han'* for *hand*) have become established variants of their respective full forms.

In addition to reduction, another diachronic effect of frequency is preservation. HF forms are more likely to survive diachronically than LF forms, and complex HF forms show a greater resistance to regularisation phenomena, like analogical levelling. A well-known example is irregular past tense forms, which tend to have a high token frequency because the LF verbs succumb over time to the pressure of regularisation and become integrated into the regular past tense pattern (e.g., German *buk* > *backte*, English *leapt* > *leaped*). Bybee (2010; p. 75) attributes this effect to what she calls *lexical strength* (which can refer to monomorphemic lexemes as well as complex constructions) as a consequence of repeated direct access: "Each use of a word or construction increases the strength of its exemplar cluster, making that word or phrase more accessible lexically" and hence more resistant to analogical change. Similarly, high token frequency can also lead to grammaticalisation and a loss of compositionality, as is the case for the English *be going to* construction. Apart from these preservation and fossilising effects, high token frequency can also lead to wider and more varied use of a structure. The higher the token frequency of a construction, the more likely it is to be used innovatively, more specifically by attracting new types. Therefore high token frequency can lead to higher type frequency (De Smet, 2016; Rohe, 2019). Likewise, the more frequent a sublexical unit (a morpheme, an articulatory gesture, etc.) is, the more it attracts future usage. Kapatsinski (2014; p. 14) calls this phenomenon "rich-get-richer positive feedback loops" and Bybee (2010; p. 53) a "self-feeding process". This tendency comes about because frequent phrases are easy to access and consequently encourage further usage (a form of long-term priming, cf. Diessel, 2017), so that their frequencies stay high or increase even further. However, frequency effects in general are often not noted "until some degree of frequency has accumulated" (Bybee, 2010; p. 18).

As has been shown, many frequency effects are related to the facilitation and strengthening of HF forms. The question nevertheless remains: which mechanisms lead to these effects? On a general level, the concept of cognitive underspecification can be taken as an explanation as to how relying on frequencies makes the processes of speech perception and production more efficient. "[H]uman cognition is biased to select contextually appropriate, high-frequency responses in conditions of underspecification" (Reason, 1992; p. 88). This automatic "schema" mode contrasts with an attentional control mode. In situations that do not require a lot of attention, a schematic action is performed that has proven successful in similar contexts.

It is still open to debate whether frequency effects arise with complex HF forms because HF patterns have stronger mental representations since *only* frequent sequences of units (phones, morphemes, or words) are stored and accessed as single units (cf., e.g., Levelt & Wheeldon, 1994; for frequent syllables being stored, also discussed in Section 8.2.1), or because their components can be assembled faster due to strengthened processing paths (as implemented in many connectionist models in the form of stronger connection strengths for HF combinations, e.g., Luce, Goldinger, Auer, et al., 2000). Concerning the holistic storage and processing of HF units, chunking in connection with exemplar models (see above) provides a natural explanation for the mechanisms involved: chunks can be stored as exemplars (Bybee, 2010; p. 38), and, since each encounter with a linguistic unit is stored in memory, the cloud of exemplars comprising this specific longer sequence becomes larger, along with the clouds for its components. In this way, the unit representation for the sequence is strengthened and later access to the whole sequence is facilitated. Bybee (1999) explicitly argues that high token frequency impacts strength of representation.

On the other hand, connectionist approaches to phonological processing hold that every encounter with a phonological pattern helps in forming the connection weights in the network.[5] Strengthening of connection weights between nodes as a function of their co-occurrence is an explanation built into connectionist models with excitatory connections between nodes of the same layer, as in PARSYN (cf. chapter 4.4.1). Both mechanisms can explain the empirical data on frequency effects

[5] In a localist network, the weights strengthened by such encounters would be those that lie between the phonemes of a phonotactic pattern (see also Section 4.4); in a distributed network, this would correspond to te strengthening of a particular activation pattern distributed over the network.

equally well and it is usually not possible to distinguish between them based on the data available (cf. Blumenthal-Dramé, 2017).

Diessel (2007) adds another point to the list. He divides the psychological mechanisms that frequency effects are based on into three categories: 1) strengthening of linguistic representations, 2) strengthening of linguistic expectations, and 3) development of automatized chunks; this reflects his background in Bybee's tradition, especially points 1 and 3. The second point, strengthening of linguistic expectations, probably incorporates a stronger conscious component. The role of expectation in linguistic processing has been studied most extensively in the syntactic domain and can be observed in the neurological P600, for example, an effect that occurs after a violation of expectations (Osterhout & Mobley, 1995). Empirical support for the role of expectancy in phonological processing comes from a study by Brown and Hildum (1956) in which two groups of native listeners had to identify English words and nonwords. The two groups had different expectations concerning phonotactic legality, and it was shown that these expectations had a significant influence on recognition rates (see Section 5.2.3).

As Jurafsky (2003; p. 36) remarks, "Probability theory is a good model of language processing [...]" under uncertainty. This influence of probability and expectations is implemented, for example, in the ART model of speech perception (cf. Section 4.4.2), in which top-down expectations from long-term memory are constantly compared against the auditory input.

In sum, frequency effects are among the most common findings in both perception and production studies, and can thus be regarded as well-established phenomena in psycholinguistic research. They follow logically from language users' exploitation of probabilistic distributions in order to reduce processing effort. Ellis (2002; p. 143) even notes that "language processing is intimately tuned to input frequency." A number of mechanisms have been described to account for frequency effects, ranging from schematisation, chunking and strengthening of representations or neural connections to stronger expectations of HF structures. It is difficult to differentiate between these explanations: it is possible that several mechanisms contribute to the frequency effects observed. Nonetheless, the underlying principle is always the same: subconscious tracking of probability distributions in the language we are exposed to. Ellis (2002; p. 148) sums it up as follows: "Type or token units, exemplar, [...] or connectionist mechanism, these are importantly different variants of figuring, but it is all counting, one way or another, and it is all unconscious." Language users are able to store enormous amounts of distributional information derived from this unconscious counting, and this knowledge guides their expectations in processing. However, researchers in usage-based linguistics stress that frequency is by no means the only factor influencing language processing and change, and that it interacts with other factors, such as analogy and information processing (cf. Diessel, 2007).

More detailed discussions of frequency effects in speech perception and production will follow in the respective chapters.

3. Consonant sequencing and consonant clusters

3.1. Introduction

In the three studies reported in this dissertation, effects of sound sequencing on speech processing are examined using the example of initial consonant clusters. Language users have clear intuitions about which phoneme sequences are more or less well-formed, and these intuitions are argued to be reflected in different degrees of processing difficulty. The question of what form phoneme sequencing knowledge takes (e.g., the degree of abstraction) and where it has its source has been debated for some time. In principle, there are two possible sources: on the one hand, sequencing knowledge could be a part of Universal Grammar or be derived from general cognitive principles or on the other hand, it could be acquired through encounters with language. Very often, the first possibility is equated with rule-based, structural knowledge, while the second is equated with statistical knowledge about concrete distributions in a particular language; neither of these connections are logically compelling.

This chapter provides a short introduction to phoneme sequencing and summarises evidence for and against both views, focusing on the concept of sonority and its implications for consonant sequencing. Furthermore, consonant clusters as units in speech processing are discussed, in addition to the presentation of the specific consonant clusters used in the experiments reported in this dissertation.

3.2. Language-specific phonotactics

Consonant cluster frequencies are discussed as *gradient phonotactics* and *phonotactic distributions* in this thesis. Since phonotactics is traditionally conceived of as absolute rules concerning the admissible positioning and sequencing of phonemes in a given language (e.g., Bußmann, 1983), it might seem peculiar to present sublexical frequency as an instantiation of phonotactics. The issue is commonly approached from a language-structural perspective. However, research has shown that phonotactic rules have a psychological component in that language users are aware of their language's phonotactics and are influenced by that knowledge during speech processing (Boll-Avetisyan, 2012). The notion of phonotactics as a gradient system in the mind of the language user, which is based on probability distributions in the language, has gained influence in recent years and has been used to describe frequency effects (Albright, 2007a; Boll-Avetisyan, 2011; Frisch, 1996; Vitevitch & Luce, 1999). The more frequent a phonotactic pattern is, the more well-formed it is judged to be by native speakers and the easier it is to process (Large et al., 1998). Phonotactics can be viewed as regularities that language users derive from phoneme distributions in the lexicon. Speakers do not consciously attempt to learn them (Ellis, 2002), but they unconsciously acquire them; and this knowledge helps them to process speech more efficiently afterwards. Throughout this dissertation, sublexical frequencies will therefore be considered as the basis of gradient phonotactics, the effects of which can be seen both in acceptability judgements and during processing. Crucially, language users make predominantly subconscious use of their phonotactic knowledge during speech perception and production processes.

Phonotactic knowledge based on specific distributions encountered in language is fundamental to usage-based phonological theory (see Section2.2). Since it is well-established that humans are sensitive to distributional statistics, it is very plausible that they derive phonotactic knowledge from the distribution of phoneme sequences in the language(s) they are exposed to. As Cutler (2012; p. 447) states, "listening to speech is, from beginning to end, tailored to a specific language". For such purposes, a mechanism for acquiring native language phonotactic distributions and the ability to use this knowledge to facilitate perception of frequent phonotactic patterns would be a very efficient means of supporting speech processing.

Studies of L1 acquisition present excellent test cases to find out if it is experience with language that shapes phoneme sequencing knowledge or if this kind of knowledge is of a more general nature, present before L1 acquisition and unaltered by it. Experimental results show that knowledge of L1 phonotactic distributions is indeed obtained during L1 acquisition. While infants' listening behaviour to high- and low-probability phonotactic sequences is the same at 6 months of age, they prefer listening to nonwords of high as opposed to low phonotactic probability around the age of 9 months (Archer & Curtin, 2011; see also Jusczyk et al., 1994). This suggests that differential behaviour to different phoneme sequences is acquired during infancy and is based on experience with native language distributions. Moreover, at 12 months, infants are able to form object-word associations when presented with phonotactically legal phoneme sequences, but do not map objects onto phoneme sequences that are phonotactically illegal in their native language (MacKenzie et al., 2012). This means that lexical acquisition is also guided by knowledge of language-specific phonotactics. Furthermore, it has been shown that even adults are sensitive to phonotactic distributions after only brief exposure to them and are influenced by this newly acquired knowledge during speech production (Dell et al., 2000; Onishi et al., 2002). Taken together, these experimental findings indicate that encounters with phonotactic distributions in the surrounding language have an impact on phonotactic knowledge and subsequent utilisation of this knowledge for lexical acquisition and speech processing.

Further support for the importance of linguistic experience for phonotactic knowledge comes from studies that compare speech processing in adults with different L1s. It has been shown that speakers of phonotactically less restrictive languages, such as Russian, are more accurate at processing typologically less preferred phoneme sequences than speakers of more restrictive languages (e.g., Davidson, 2011). Knowledge of a less restrictive phonotactic system can even reduce perceptual illusions (Carlson et al., 2016; Cohen et al., 1967). This indicates that native language distributions shape our phoneme sequencing knowledge. If universal well-formedness principles were the main source of this knowledge, speakers of different languages should show the same preferences and effects. However, there are some indications that universal preferences exist; a potential underlying principle will be presented in the next section.

3.3. Sonority

The phonological concept of sonority is traditionally used to account for cross-linguistic distributions of consonant sequences, as well as their diachronic developments. According to some researchers, sonority is a part of Universal Grammar (e.g., Clements, 1990). In sonority theory, a sonority value on a scale or hierarchy, which corresponds roughly to the segment's "loudness relative to that of other sounds with the same length, stress, and pitch" (Ladefoged, 1975; p. 219) can be assigned to each class of phonemes—and in some cases also specific phonemes within these classes. This means that the sonority value is an inherent property of the phoneme, independent of the context in which it appears. The sonority values determine the optimal order of those phonemes in a syllable. The closer a syllable is to the optimal sequencing, the more common it is cross-linguistically. Though originally used to explain purely linguistic phenomena, sonority sequencing has increasingly been applied to the psycholinguistic domain. It has been found to play a role for intuitions about well-formedness as measured by acceptability ratings (Albright, 2007b), perceptual illusions in illegal sequences (Berent et al., 2007), order of acquisition in L1 and L2 (Broselow & Finer, 1991; Hamza et al., 2018), as well as production difficulties in populations with speech impairment (Miozzo & Buchwald, 2013), although none of these findings is undisputed. Overviews of the relevant results regarding speech perception and production will be given in the respective chapters.

First mentions of sonority and its structuring regularities date back as far as and Whitney (1874)Sievers (1897), who described a continuum of different consonant classes with respect to the openness of the vocal tract and referred to the syllable as having a "crescendo–diminuendo" structure (p. 293). Particularly in the last ten to fifteen years, interest in sonority has grown with a near-exponential rise in number of studies (for an overview, see Parker, 2017). One of the questions debated is whether the sonority hierarchy is universal or language-specific. Most researchers adhere to one universal hierarchy for reasons of explanatory power, but differing tendencies in individual languages challenge this view. Therefore, some minor degree of language-specific variation has been suggested (e.g., Parker, 2002), an approach that has been rejected by others (Clements, 1990). Several variants of sonority hierarchies have been proposed, which vary in granularity. The most common hierarchy (proposed, for example, by Clements, 1990) is displayed in (1).

(1) vowels > glides > liquids > nasals > obstruents

Some researchers subdivide obstruents into fricatives and stops (e.g., Zhao & Berent, 2016) or fricatives, affricates, and stops (Ulbrich et al., 2016), while some differentiate between voiced and voiceless obstruents, as well as rhotics and laterals or different types of vowels (e.g., Vennemann, 2011). The most fine-grained sonority hierarchy—

consisting of 16 different sonority classes—has been postulated by Parker (2002). He resolves the problem of universality and individual variation by claiming that this fine-grained ranking is universal but individual languages collapse several of the classes in different ways. For the purposes of the present study, the hierarchy proposed by Berent et al. (2007) and Sievers (1897), as displayed in (2), has been adopted because it allows for the investigation of potential effects on a fine-grained level (differentiating, for example, between /pl/ and /fl/ or /ʃt/ and /tʃ/); this degree of granularity has proven fruitful in psycholinguistic investigations (e.g., Miozzo & Buchwald, 2013; Ulbrich et al., 2016). Sonority values are given here for illustrative purposes. Note, however, that these values do not represent any meaningful units; rather, it is the difference between them that matters in sonority theory.

(2) vowels > glides > liquids > nasals > fricatives > stops
6 5 4 3 2 1

3.3.1. Correlates of sonority

There are two streams in sonority research: phonological accounts treat sonority as a structural phenomenon that forms the basis of phonological processes, while phonetic accounts have attempted to identify phonetic correlates to sonority (see Miozzo and Buchwald, 2013, for a juxtaposition of the two traditions). Despite its long tradition in phonology, however, sonority is notoriously ill-defined phonetically. Various proposals have related it to a phoneme's acoustic or articulatory specifications. Lindblom (1983), among others, postulated that sonority corresponds to the position of the jaw (determining the openness of the oral cavity), which he ultimately attributed to a consonants' coarticulation propensity, thereby motivating it from adaptations to the motor synergy constraints of speech production.[1] However, he uses a sonority hierarchy (developed by Elert (1970) to account for Swedish phonotactics) that places /s/ and /ɧ/ at the extreme end of the scale behind stops; thus the close correlation is not between jaw position and the standard sonority scale. This placement of /s/ also circumvents accounting for common violations to sonority sequencing presented by sibilant-initial sequences (see below).

On the acoustic side, popular candidates for sonority correlates have been intensity or acoustic energy (Heffner, 1969), formant structure (Clements, 2009; Ladefoged, 1997), and voicing (Yavaş, 2003). All have proven to be problematic, however, and Clements (1990; p. 287) concludes that "sonority remains an ill-defined, if not mysterious concept in many respects". As a result, many scholars, Clements included, have claimed that it cannot be equated with any single acoustic, auditory, or articulatory feature. Instead, it is assumed to be a composite property that correlates with various phonetic properties to a certain degree—all of them contributing to relatively higher sonority (Clements, 1990; Nathan, 1996). In recent years, the search for physical correlates of sonority has been taken up again. The most exhaustive study is Parker's (2002) dissertation, which found that sonority is correlated (in decreasing order) with intensity, intraoral air pressure, F1 frequency, total air flow, and duration. This led him to conclude that "sonority can and should be defined as a function of intensity" (Parker, 2002; p. 242). Nevertheless, he concedes that "the sonority hierarchy needed for actual phonological analyses may differ in minor yet principled ways from the scale which emerges from my phonetic results". (Parker, 2002; p. 231). All in all, he interprets these findings as support for the physical reality of the sonority hierarchy.

[1] Also Code and Ball (1994; p. 265) consider it likely that sonority "is simply [...] a non causal consequence of neurophysiology and the mechanico-inertial constraints of the speech production mechanism."

Among phonologists, there is a higher level of consensus concerning the basis of sonority. It is widely agreed that sonority values can be derived from specifications for the major class features of phonological theory. The features [syllabic], [vocalic], [approximant], and [sonorant] are all correlated with sonority to a certain degree. The more of those features a phoneme is positively specified for, the higher its sonority value will be (Stenneken et al., 2005).[2] According to Clements (1990), the phonetic correlates of sonority are exactly those of the major class features, which have in common the fact that they all contribute to the overall perceptibility of a phoneme. Approximately 20 years later, however, he reduces the phonological basis of sonority primarily to the feature [+sonorant]. He argues for perceived resonance, whose main property is prominent formant peaks, as the phonetic correlate of sonority and points out that these formant peaks are exactly what defines the feature [+sonorant]. "The sonority scale then corresponds to the degree to which a given segment possesses the characteristic properties of [+sonorant] sounds" (Clements, 2009; p. 5). This property decreases from vowels and semivowels via liquids and nasals to obstruents.

To sum up, there are a number of both phonological and phonetic factors that contribute to the notion of sonority, but opinions vary on whether all of these factors are needed to define it or whether only one is close enough to be considered an "absolute" correlate.

To this day, criticism as to the utility and adequacy of sonority remains (e.g., Everett, 2016; J. J. Ohala, 1992; J. J. Ohala & Kawasaki, 1984), with objections including that the concept is circular and devoid of any empirical basis (Everett, 2016), and that it offers "no principled explanation" (J. J. Ohala and Kawasaki, 1984; p. 122) for observations of syllable structure. Accordingly, suggestions have been made to replace it

[2] For a discussion of the redundancy of having both binary major class features and the multi-valued feature of sonority in phonological theory, see Clements (1990).

with more independently motivated principles (e.g., cue robustness for speech perception, cf. Henke et al., 2012).

Clements (2009; p. 1) notes, "the ultimate justification for such concepts depends on their success in bringing order to a vast array of seemingly disparate facts"—facts that cannot be accounted for by other factors, it may be added.

3.3.2. Principles in sonority

The most important syllable structuring principle in sonority theory is the Sonority Sequencing Principle (SSP; also called Sonority Sequencing Generalisation, SSG). A definition, as stated by Clements (2009), is as follows:

Sonority Sequencing Principle. *Segments are syllabified in such a way that sonority increases from the margin to the peak.*

This essentially means that, in onset clusters, the second member should have a higher sonority value than the first one, while the opposite is true in coda clusters. Clusters in which both consonants have the same sonority value (plateau clusters) or in which the outermost consonant has a higher sonority value than the one adjacent to the syllable peak are considered violations of the SSP. According to Clements (1990), the SSP holds at an abstract level prior to the application of phonetic realisation rules (at the level of lexical phonology). It "expresses a strong cross-linguistic tendency" (Clements, 1990; p. 301). Consonant clusters conforming to the SSP are more common across the world's languages, and in many languages, they are the only ones allowed (Clements, 1990). Beyond this absolute principle, which divides syllables into marked (those that violate the principle) and unmarked ones (those that conform to it), there is an additional, weaker, preference law for syllables, which expresses the relative complexity or preferability of a syllable: the Sonority Dispersion Principle (SDP).[3]

Sonority Dispersion Principle (quoted from Clements 1990, p. 303). *The simplest syllable is one with the maximal [...] rise in sonority at the beginning and the minimal drop in sonority [...] at the end. Syllables are increasingly complex to the extent that they depart from this preferred profile.*

Thus the steeper the sonority rise at the beginning of a syllable, the more common the syllable is cross-linguistically (Berent, 2016).

In addition to these universal sonority principles, every language is said to have a specified Minimal Sonority Distance (MSD) between adjacent onset phonemes (J. Harris, 1983). The sonority distance between two phonemes is calculated by subtracting the sonority value of the first consonant from the sonority value of the second consonant. For example, the sonority distance in the cluster /kl/ is 3 (4–1), and in the cluster /ʃt/ it is -1 (1–2). In Spanish, for example, the necessary minimal distance between two onset consonants is 2 on the basis of the scale given in (2), which bans obstruent–nasal and nasal–liquid onsets since they have a sonority distance of 1. Conversely, in Korean, the necessary minimal distance is 4, which allows only for stop–glide clusters, and in Russian it is 0, which licenses even plateaus. As has been shown, the SSP is supposed to be universal, whereas the MSD is language-specific. Both combine to set the sonority rules for syllables in a given language. These principles were already referred to by Whitney (1874), albeit not under their current names.

[3] From the SDP follows the so-called Syllable Contact Law, which requires sonority to fall across syllable boundaries. As the study here is concerned with the goodness of specific onsets that occur in various contact contexts in everyday speech, this principle is not relevant to the present study and will not be discussed further.

3.3.3. Exceptions to sonority sequencing

A number of researchers (e.g., Ott et al., 2006) have argued that the sonority hierarchy is not valid for syllable-initial clusters with a sibilant in C1 position[4] since a subgroup of them, /s/ + stop clusters, such as English *stand*, present very common violations to the SSP. To deal with this problem, several options have been proposed, which range from extra-syllabicity (J. Harris, 1994) to even representation as a single segment (Selkirk, 1982). Already Sievers (1897) introduced the notion of a *secondary syllable* ("Nebensilbe"), "a unit which counted as a syllable for the purposes of the SSP but not for linguistic rules" (Clements, 1990), to deal with sC clusters. Obviously, such a compromise is neither elegant nor satisfying. For a full discussion of sC clusters, see Goad (2011). Here, it is sufficient to state that sC clusters present prominent exceptions to sonority sequencing and that there is currently no consensus on how to deal with them.

It should also be kept in mind that, according to Clements (1990), the domain of sonority is core syllabification, that is, initial syllabification. Rules that apply to the periphery of syllabification can introduce more complex syllable types than those produced during initial syllabification. Laeufer (1995), by contrast, interprets her findings concerning fast speech, in which the SSP is still obeyed, as evidence that the SSP holds at later levels of representation as well.

Similarly, Parker (2017) draws attention to the importance of studying the phonetic realisation of phonemes that constitute apparent violations of the SSP in order to determine whether such clusters are indeed tautosyllabic.[5]

[4] For convenience, they will henceforth be referred to as sC clusters, even though they surface as /ʃ/C in German.

[5] With respect to German, an EPG study by Bombien et al. (2010) showed that /sk/ clusters are produced with less overlap than /kl/, but with more overlap than /kn/. Pouplier (2012) found that C2 moves towards the vowel in /sk/ clusters, which suggests that the /s/ might indeed not be part of the onset.

In conclusion, it can be said that sonority continues to be employed as an explicans for a wide array of linguistic and psycholinguistic phenomena, but nevertheless remains controversial on a phonetic level. It also fails to explain the cross-linguistically common violations to syllable structure presented by initial sC clusters. Any sonority account that is consistent with empirical data either treats sibilants at syllable margins as extrasyllabic or places them outside of their natural class on the sonority scale. However, in a recent metastudy of 264 experimental investigations on the phonetic, phonological, cognitive, and neurophysiological manifestations of sonority, Parker (2017) concludes that, although evidence is mixed, there is "moderate support for the linguistic relevance of sonority".[6] How sonority principles behave in relation to other factors in speech perception and production is still open to debate. Following the interpretation of their aphasic data that sonority and frequency effects in speech production are based on different mechanisms, Miozzo and Buchwald (2013; p. 298) conclude that "further evidence is needed before reaching firmer conclusions on the relationship between sonority and frequency, not only because it is desirable to obtain additional converging evidence but also because the complexity of such [a] relationship requires a systematic investigation." This dissertation is aimed at shedding more light on this intricate relationship. For a very detailed discussion of sonority, its potential physical correlates and exceptions, see Clements (1990).

3.3.4. Sonority in German

A sonority analysis of the German CELEX database by Stenneken et al. (2005) revealed that the syllable type that is most preferred on a sonority basis (obstruent–vowel, because of the steep sonority rise in the onset and the minimal decline in the offset) is the most frequent syllable

[6] This can be quantified as 57% of the studies reviewed supporting accounts of sonority or, at a minimum, are consistent with them.

type in the German language, both in terms of type and token frequencies. Moreover, even the other—less frequent—syllable types show a "tendency towards a maximum sonority contrast" (Stenneken et al., 2005; p. 289) in syllable onset position (e.g., obstruent–liquid–vowel being more frequent than obstruent–nasal–vowel). When it comes to initial consonant clusters, Orzechowska and Wiese (2015; p. 441) note that "German tends to form clusters obeying the SSG"; they found that almost 70% of onset clusters have a rising sonority profile. However, some of the most common clusters are sC clusters and thus arguably violate the SSP. For example, the single most frequent initial cluster in German is /ʃt/. This discrepancy is certainly striking. Interestingly, in their hierarchical cluster analysis (the term *cluster* being used in its statistical sense) Orzechowska and Wiese (2015) found that the resulting division into the two major groups of consonant clusters (in the phonological sense) is determined by their adherence to the SSP.

For an in-depth analysis of German initial consonant clusters with respect to sonority (inter alia) from a structural perspective, the reader is referred to Orzechowska and Wiese (2015).

3.4. Alternatives to sonority

3.4.1. Net Auditory Distance (NAD)

Net Auditory Distance (NAD) has been suggested as a (perceptual-) salience-based explanation for universal preferences for certain consonant clusters since it makes finer predictions than the SSP (Dziubalska-Kołaczyk, 2014). This section will give a very brief overview of the underlying reasoning, the calculation of NAD values for consonant clusters, and the preference predictions for initial clusters made by NAD theory as a basis for comparison with sonority and the SSP.

NAD is a principle stemming from Beats-and-Binding Phonology (B&B Phonology, Dziubalska-Kołaczyk, 2001, 2009), a branch of Natural Phonology that is grounded in phonetics and replaces the traditional notion of the syllable with acoustically determined alternating sequences of beats (mostly vowels) and non-beats (consonants). In general, consonant clusters tend to be avoided due to the universal preference for non-beat–beat structures (i.e., CV structures). However, different consonant clusters vary in their degree of well-formedness, which can be expressed by the NAD values of the phonemes in the sequence. The underlying idea is that perception is grounded in contrasts (*modulation*, see also J. J. Ohala, 1992), so that a succession of two segments is more easily perceptible if there is a certain degree of auditory contrast between them, with CV structures exhibiting the maximal contrast and hence the best perceptibility. This contrast can be expressed as their NAD, the "net auditory impression of a distance between consecutive segments in a sequence" (Dziubalska-Kołaczyk, 2019; p. 112). The NAD between two adjacent phonemes is a measure of the distance between them in terms of manner of articulation (MOA) and place of articulation (POA). Voicing is not taken into consideration because it is a redundant feature for several phoneme classes. Each POA and MOA is assigned a specific value so that a simple numeric difference between two phonemes can be calculated separately for POA and MOA. The resulting values are then summed in order to produce a NAD value for the phonemes concerned (see calculation in (4) below). Table 3.1 shows the values given for the German consonant inventory.

The preferability of consonant clusters then "reflects the strength of the contrasts between cluster constituents and how they compare to the contrasts between cluster constituents and neighbouring vowels" (Dziubalska-Kołaczyk, 2019; p. 112). For syllable-initial consonant clusters, the preference is for the NAD between the two consonants to be greater or the same as the NAD between C2 and the following vowel (see equation in 3). This means, the higher the NAD between the two consonants in relation to the distance between C2 and the following vowel, the more well-formed the sequence is.

obstruent			sonorant				vowel		
stop	affricate	fricative	nasal	lateral	rhotic	glide			
5.0	4.5	4.0	3.0	2.5	2.0	1.0	0		
p b			m				1.0	bilabial	labial
	pf	f v					1.5	labio-dent.	
t d	ts	s z	n	l			2.0	alveolar	coronal
		ʃ ʒ					2.5	post-alv.	
		ʝ				j	3.0	palatal	dorsal
k g			ŋ				3.3	velar	
					ʁ		3.6	uvular	
		h					5.0		glottal

Table 3.1.: NAD values of German consonants
Table adapted from the NAD Phonotactic Calculator (http://wa.amu.edu.pl/nadcalc/, Dziubalska-Kołaczyk et al., n.d.)

(3) $\text{NAD (C1,C2)} \geq \text{NAD (C2,V)}$

(4) $\text{NAD (C1C2)} = |(\text{MOA1} - \text{MOA2})| + |(\text{POA1} - \text{POA2})|$

For instance, the sequence /prV/ is preferable to the sequence /plV/ because the distance between /p/ and /r/ is greater than the distance between /p/ and /l/; and so is the so-called NAD product, which takes into account the transition to the following vowel. The calculations in (5) and (6) demonstrate this.

(5) $\text{NAD (C1C2)} = |(\text{MOA1} - \text{MOA2})| + |(\text{POA1} - \text{POA2})|$
$\text{NAD /pr/} = |(5 - 2)| + |(1 - 3.6)| = |3| + |-2.6| = 5.6$

$\text{NAD (C2V)} = |(\text{MOA1} - \text{MOA2})|$
$\text{NAD /rV}^{7}\text{/} = |2 - 0| = 2$

[7] Ideally, vowels would be differentiated according to their quality; however, all vowels are assigned a MOA value of 0 in the NAD tables provided by Dziubalska-Kołaczyk (2014) and Dziubalska-Kołaczyk et al. (n.d.).

NAD (C1C2) – NAD (C2V) = 3.6

(6) NAD (C1C2) = |(MOA1 – MOA2)| + |(POA1 – POA2)|
NAD /pl/ = |(5 – 2.5)| + |(1 – 2)| = |2.5| + |−1| = 3.5

NAD (C2V) = |(MOA1 – MOA2)|
NAD /rV/ = |2.5 – 0| = 2.5

NAD (C1C2) – NAD (C2V) = 1

As can be seen, NAD makes finer distinctions than sonority because it considers the difference in POA in addition to the difference in MOA. However, it faces the same problem concerning sibilant–stop clusters as sonority does: these clusters are dispreferred in terms of NAD as well.

3.4.2. Generalised phonotactics

Language-specific but generalised phonotactics have, in particular, been suggested as an alternative—or in addition—to universal, innate biases in order to account for preference hierarchies among unattested sound sequences (Albright, 2009). This includes phonotactic rules for which phoneme classes can occur in a sequence, as well as generalised consonant cluster frequencies (cf. Section 2.2).

Language users have preferences for some illegal consonant clusters over others. In addition to these conscious preferences determined in metalinguistic judgement tasks, there is also experimental evidence for direct effects in processing structurally different consonant clusters of zero frequency. English listeners show perceptual illusions in the perception of initial [dl] but not [bw], although both are equally unattested in English (Moreton, 2002). Moreton (2002) explains listeners' perceptual repairs such as *[dl] > [gl] (the closest legal alternative in the listeners' language) as reflecting a generalised ban of [coronal][coronal] sequences in language users' grammars. He argues that these feature-based generalisations can explain this perceptual phonotactic bias, which segment-based phonotactic rules or consonant cluster frequencies cannot account for.

Furthermore, in artificial grammar learning experiments, participants learned generalisations over natural classes even before they acquired segment-specific phonotactics (Linzen & Gallagher, 2014). In a similar fashion, generalisation over gradient phonotactics (i.e., biphone frequency distributions) has been shown to aid speech segmentation (Adriaans & Kager, 2010). Taken together, this indicates that phonotactic knowledge 1) does indeed involve generalisation over individual segments and 2) is used both for judgements about linguistic items and during speech processing. Daland et al. (2011) claim that all that is needed to make seemingly SSP-based generalisations regarding onset clusters purely on the basis of the native lexicon is "a sufficiently rich representation of phonological context", for example, the ability to differentiate between onset, coda and heterosyllabic consonant clusters, and a featural representation that allows for the capture of sonority.

Importantly, however, the fact that phonotactic distributions can be generalised over does not necessarily exclude phonotactic knowledge about the well-formedness of specific phoneme sequences. In most models of generalised phonotactics, rules can take on different degrees of generality, and those concerning specific phonemes present the highest degree of specificity. In cases where the specific rules are in conflict with more general ones, they are likely to win them over and establish an "exception" from the general rule. In a comparison of different computational models, Daland et al. (2011) found that models based on segments best predict human ratings of attested clusters, while models that make featural generalisations are best at predicting ratings of unattested clusters. This suggests that language users only resort to generalised probabilities when segmental probabilities are unavailable.

3.5. Consonant clusters

3.5.1. Consonant clusters as units

This section addresses the object of the present study, namely consonant clusters, and the question of how they are defined and what makes them an interesting object of study.

What stands open to debate in the present dissertation is how experience and structure influence language processing. In order to investigate this question, adequate units have to be used; it will be argued that consonant clusters represent such units. Conversely, finding effects of consonant cluster frequencies in the studies will also support the notion of consonant clusters as units in speech processing.[8] As Arnon (2015; p. 274) noted, "Frequency effects [...] are interesting because they reveal something about the learning mechanisms and *units* [emphasis added] used in language learning".

From a language-structural perspective, a consonant cluster is a sequence of several consonants without an intervening vowel. A distinction is made between phonotactic and morphonotactic clusters. Phonotactic consonant clusters occur within morphemes, whereas morphonotactic clusters emerge as a result of morpheme sequences and contain a morpheme boundary, for example, English /md/ as in *seem+ed* (cf. Korecky-Kröll et al., 2014). The present study is limited to phonotactic clusters. Moreover, it investigates only tautosyllabic clusters, more specifically, clusters in syllable onset position. Consonant clusters dif-

[8] Note that the presence of frequency effects is not unambiguous evidence for the reality of a unit, though. What seems like the effect of a unit of a specific size can, in connectionist models, often be derived without explicit reference to the unit. Wade et al. (2010) have shown this for apparent syllable frequency effects (cf. also Dell et al., 1993a).

fer from affricates in that the elements of a cluster are distinct segments of their own, whereas affricates are complex segments that consist of several parts. While this distinction is a phonological one, it is also relevant for psycholinguistics. For example, affricates are acquired before consonant clusters (Lléo & Prinz, 1997). Phonetic differences between affricates and consonant clusters are highly debated but concern primarily temporal properties, like fricative rise time. A full discussion of them is beyond the scope of this dissertation; detailed analyses can be found in Griffen (1981) and Reetz and Jongman (2009).

It is here hypothesised that consonant clusters are a processing unit between the single phoneme and the syllable, tied together by entrenchment just like frequent multi-word sequences. In a language like German, which allows a large number of both initial and final consonant clusters, processing them as units would increase efficiency (for example, by reducing phonological working memory load, cf. Segawa et al., 2019) compared to recognition and later combination of the constituent phonemes in perception or the separate planning and combining of separate articulation plans in production. The idea that consonant clusters behave as holistic units is by no means new (Bond, 1971; Cutler et al., 1987; MacKay, 1972). However, experimental evidence is inconclusive: Bond (1971), Cutler et al. (1987), Hallé et al. (1998), and Segawa et al. (2019) argue in favour of this position, for example, while Newton (1972) and Ziegler et al. (2008; for phonetic planning in apraxic production) argue against it. Analyses of possible frequency effects, as undertaken in this study, will shed further light on the issue since frequency effects have been reported for linguistic units of different sizes (e.g., phonemes, syllables, words, phrases) which have proven relevant to speech perception.

Several psycholinguistic studies suggest that tautosyllabic consonant clusters are psychological real units of speech processing. Consonant clusters are broken up by phonological speech errors in only 6% of cases, while 22% would be expected by chance (MacKay, 1972). When asked to name words beginning with certain letters, subjects show shorter production latencies when presented with consonant clusters than with either singleton consonants or CV sequences (Claxton, 1974). A similar situation is found in speech perception: in phoneme monitoring, subjects are faster in detecting a target consonant when the cluster it occurs in is the same in the target word as in the model given for the monitoring task ("Listen for a *b* as in *blue*") than when the target word begins with a target phoneme that is part of another cluster or a simple onset (e.g., *brevity* or *barn*; Cutler et al., 1987). All of the above are an indication that consonant clusters are stored and processed as holistic units. Furthermore, Bond (1971) interprets the time course of consonant cluster recognition as indicative of consonant clusters being perceived and processed as one unit.[9]

On the other hand, Berg (1989) finds varying levels of cluster cohesiveness in speech errors and concludes that "[t]hey may act as a unit on one occasion but split on another" (Berg, 1989; p. 258). He attributes this variable behaviour to sonority relations between the consonants of a cluster, claiming that Cl clusters are more cohesive than Cr clusters and that non-liquids are "even more closely tied to C1" (Berg, 1989; p. 261). However, when one takes a closer look at the individual cases he lists, an alternative explanation for the varying behaviour of consonant clusters can be formed on the basis their frequency. While the HF clusters /ʃt/ and /ʃp/ act as a unit in almost all cases, LF /ʃr/ is consistently split up by speech errors. Medium-frequency clusters like /fr/ and /kl/ behave as a unit in approximately half of all cases.

This differential behaviour can easily be explained in terms of chunking. As laid out in Section 2.1, usage-based linguistics holds that units

[9] Most evidence for consonant clusters acting as a unit comes from onset clusters (Levelt et al., 1999). It has also been suggested that word onsets are cohesive units in themselves (Treiman, 1986). The debate about the special status of onsets is extensive (see, for example, Connine et al., 1993; Gow et al., 1996) but will not be pursued any further here.

that are repeatedly used together eventually fuse together (Bybee, 2002), they "face a pressure to become automatized" (Ibbotson, 2013; p. 2). While chunking—the storage and processing as holistic units—has mostly been demonstrated at the syntactic level, it is equally advantageous on the phonological level. Since the same principles and mechanisms are supposed to apply to all linguistic levels of processing (Bybee & McClelland, 2005), it is reasonable to assume that phonemes are chunked into larger phonological units. Recall that entrenchment is a gradual process that depends on frequency of use. It is therefore expected that HF clusters, like /ʃt/, behave most unit-like in speech processing, while LF clusters, like /ks/, do not.[10]

What further makes two-consonant clusters a likely candidate for units in speech processing (and their frequencies a relevant factor) is the fact that the prediction of diphone statistics from phoneme frequencies is not possible. Triphone statistics, on the other hand, can be inferred from diphone statistics. Pierrehumbert (2003; p. 146) concludes, therefore, that "diphone statistics can be, and must be learned".

3.5.2. Consonant clusters used in the present study

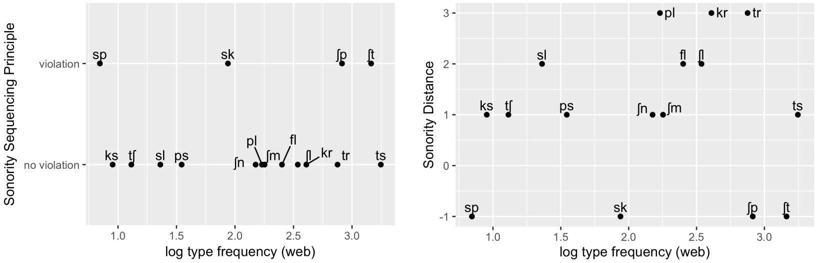
German allows 56 different initial consonant clusters (Orzechowska & Wiese, 2015). Of these, the following 16 consonant clusters have been chosen as test clusters for all three studies reported here: /ts/, /ʃt/, /ʃp/, /tr/[11], /kr/, /ʃl/, /fl/, /ʃm/, /pl/, /ʃn/, /sk/, /ps/, /sl/, /tʃ/, /ks/, and /sp/.

The selection criterium was a high number of "minimal pairs", meaning two clusters that differ in only one broad articulatory feature (man-

[10] Note that the existence of a representation of the cluster as a whole does not exclude representations of its components and the possibility that their properties (e.g., frequencies) play a role as well. It is widely assumed that there are multiple levels of representation (see, e.g., Blumenthal-Dramé, 2012), all of which can be utilised during processing.

[11] The rhotic is realised as a velar fricative or trill in Standard German and in the spoken stimuli. Since there is no phonemic distinction between [r] and [ʁ], it will be encoded as /r/ in phonemic transcriptions for better readability.

ner of articulation, place of articulation, voicing) or which consist of the same two consonants in reversed order (e.g., /sp/ and /ps/), with diverging frequency and/or sonority values. This was done in order to be able to separate the effects of frequency and sonority more reliably from articulatory and acoustic effects, and to obtain useful pairings for the stimuli in the production experiment. As Figure 3.1a shows, 25% of the clusters violate the SSP and both they and the SSP-conforming clusters are distributed over a wide range of frequency values. Figure 3.1b gives a more detailed picture since it provides the sonority distance values for the individual clusters.



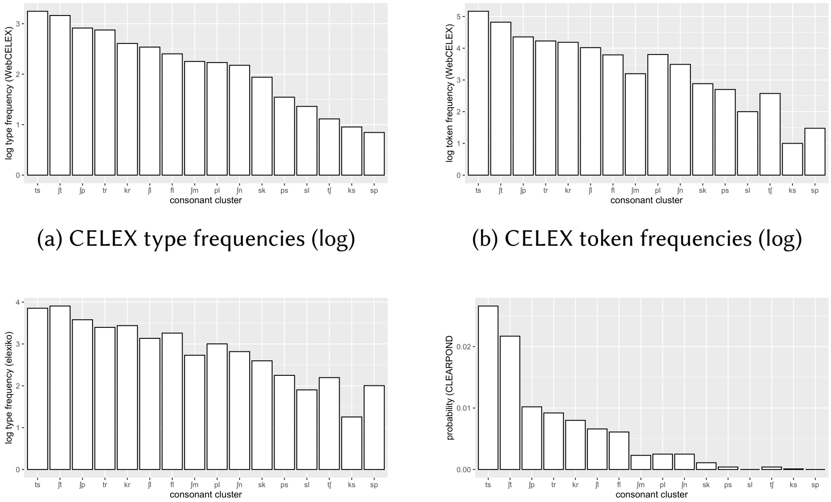
(a) Frequencies and SSP conformity

(b) Frequencies and sonority distance

Figure 3.1.: The 16 test clusters used in the experiments

The inclusion of /ts/ into the set calls for further explanation. The complex onset /t͡s/ is an affricate in German phonology. Nonetheless, it has been included in the set of *consonant clusters*. This was done because of its strong structural similarity to the true clusters /ps/ and /ks/, from which it differs greatly in terms of frequency. The choice can also help in investigating whether a difference in status leads to a processing advantage for affricates as compared to true clusters. If clusters display varying degrees of cohesiveness—as Berg (1989) suggested—, this difference should not cause a dichotomous effect. Its special status will be kept in mind when interpreting the results of the studies.

The frequencies reported in the experiments are mainly CELEX frequencies derived from WebCELEX (http://celex.mpi.nl/) by querying lemmas in the Mannheim part of the corpus for syllables beginning with the respective consonant clusters (annotated in SAMPA). To obtain type frequencies, the resulting lemmas were counted. The lemma database was used, as lemmas are used for most psycholinguistic experiments (Hofmann et al., 2007) and word forms are not relevant for the present studies. Within the entries of the database, phonological transcriptions were used because some of the onset clusters can be spelled in different ways and some spellings can represent several German clusters (e.g., <sp> can represent [ʃp] or [sp]). To obtain token frequencies, the corpus frequencies of these lemmas were summed. There has been a considerable amount of criticism concerning CELEX frequencies and their continued use because other frequency sources, such as television subtitles, have been shown to more reliably predict behaviour in psycholinguistic experiments (e.g., Brysbaert et al., 2011). While some of the criticism concerning CELEX is justified, there does not seem to be a more reliable source of subsyllabic frequencies for the present enterprise. CLEARPOND (Marian et al., 2012), a television-subtitle-based corpus, which serves as an excellent frequency measure for English, uses faulty phonological transcriptions for several phonemes in German. Therefore, the calculations of lexical neighbourhoods are not reliable, and it was considered important to base sublexical frequencies and neighbourhood measures on the same source. Another good source for type frequencies is the elexiko online dictionary provided by the Leibniz-Institut für Deutsche Sprache ("elexiko," 2003). However, there is no English equivalent for the German values derived from it, which means that for the L2 listening experiment (Chapter 6) German and English frequencies could not be compared on the basis of equivalent sources. This is important for consistency and reliable results, however. Likewise, using elexiko frequencies in the L1 experiments and CELEX frequencies in the L2 experiment would have reduced con-



(a) CELEX type frequencies (log)

(b) CELEX token frequencies (log)

(c) elexiko type frequencies (log)

(d) CLEARPOND-based probabilities (token)

Figure 3.2.: Frequencies of the 16 test clusters

sistency and comparability among the experiments. Although CELEX type frequencies served as the main source for consonant cluster frequencies, frequencies from the other sources were additionally used whenever possible and results compared. As Hofmann et al. (2007) noted, a contribution to the type vs. token controversy that takes into account both measures and compares them is needed. This is what is done here. For an overview of the frequency values for test clusters according to each of these sources, see Figure 3.2 (a)–(d).

4. Prelexical speech perception

General background and models

4.1. Introduction

Speech perception is an extremely complex process that encompasses everything from the extraction of the speech signal from the acoustic environment to the mapping between acoustic form and linguistic meaning. In spite of its complexity, it is mostly—at least in the case of L1 perception—achieved without any noticeable difficulty. The principles underlying speech perception have been investigated for many decades. Yet many of the sub-processes and influencing factors are still not fully understood to this day. Many models of speech perception have been proposed, and there are ongoing discussions as to which best accounts for the empirical findings of speech perception research.

This chapter provides the theoretical foundation for the perception experiments in Chapters 5 and 6. Before experimentally investigating the effects of frequency and sonority sequencing on speech perception in the next two chapters, this chapter describes the general background against which they must be regarded. It first gives an overview of prelexical processes in speech perception by describing the steps relevant to the listening experiments and the general factors identified that influence consonant and consonant cluster identification. It then presents two connectionist models, which can account for the mechanisms at work during prelexical speech processing: 1) the domain-general Adaptive Resonance Theory (ART) of sensory perception and information processing and 2) PARSYN, a model of spoken word recognition.

4.2. Issues and steps in prelexical speech perception

Speech perception is assumed to be accomplished in several successive (and partly overlapping) steps (e.g., Cutler & Clifton, 2000; McQueen & Cutler, 2013; Pisoni & Sawusch, 1975). The processes of speech perception that are executed prior to and for the purpose of accessing lexical items are referred to as *prelexical processes*. The aim of this study is to shed light on prelexical processes and sublexical components of speech perception. In order to recognise words, listeners must have a sound representation that they can compare to entries in the mental lexicon in a matching process. Whether this representation consists of abstract units—as proposed by abstractionist models—or is a detailed acoustic image that includes all indexical information—as advocated by episodic models—, or both is still subject to very lively debate (e.g., Ernestus, 2014; McQueen et al., 2006; Pierrehumbert, 2001).

Some researchers assume a direct mapping of the acoustic–phonetic signal onto lexical–semantic representations without any intervening prelexical recognition processes (e.g., Klatt, 1979). If this were the case, the distributional characteristics of sublexical units, such as phones or syllables, should not lead to any observable effects on perception accuracy or response time. However, a group of researchers centred around Paul Luce has consistently found facilitative effects of phoneme probability on processing latencies (to be discussed in Section 5.2.4). Luce and Large (2001) thus conclude that word recognition models including the recognition of sublexical units are more realistic. In fact, since there are empirical findings in favour of the use of both abstract and episodic information (for a short overview see McQueen & Cutler, 2010), it is highly likely that both are used during speech perception in some way.

Based on Luce's findings about effects on the sublexical level, the present work assumes that abstract linguistic units, such as phones, play a role in auditory speech processing and that their frequencies can thus have an effect on their perceptibility. Nevertheless, this supposition does not mean that the storage of detailed acoustic information is rejected. Given that the inclusion of fine phonetic detail in mental representations is beyond the scope of this thesis, it will not be discussed further.

The nature and size of the potential abstract unit is similarly the subject of fierce debate—and has been for decades—as the question of whether it exists at all. The most influential propositions include features (Eimas & Corbit, 1973; Mesgarani et al., 2014), articulatory gestures (Liberman et al., 1967; Studdert-Kennedy, 1998), allophones (Wickelgren, 1969), phonemes (Decoene, 1993; Foss & Blank, 1980; Kazanina et al., 2017), demisyllables (Samuel, 1989), and syllables (Massaro, 1972; Mehler, 1981). The existence of all aforementioned proposed units has been supported by empirical findings, and so it is conceivable that all of them play a role in perception processes. In fact, the inference that there might not be a single "basic unit" of speech perception at all but, rather, different units are used at different levels of perception is about as old as the suggestions concerning the various units themselves:

> It should be obvious that the size of the processing or structural unit will vary as a function of the level of processing in the linguistic system. [...] To argue that there is one basic unit in speech perception is to acknowledge that language exists primarily in one form rather than another. However, the major fact about human language is that it exists in many forms, most of which are inaccessible to

> conscious inspection. (Pisoni and Sawusch, 1975; pp. 18–19)

The question of which unit figures prominently during processing depends on task demands, namely the demands of the specific listening situation (McQueen & Cutler, 2010). Throughout this chapter, it will be assumed that the recognition of phoneme- and biphone-sized units (i.e., consonants and consonant clusters, respectively) is involved in prelexical speech perception.

Very broadly speaking, the processing stages executed prior to (and partly overlapping with) lexical activation include the extraction of the speech stream from its acoustic environment, arguably the normalisation of the extracted signal into abstract units, and segmentation into candidates for lexical access. For an excellent overview of the different processes involved in speech perception (including prelexical processes), see Cutler and Clifton (2000).

The prelexical process that is the subject of this study is the recognition of sequences of phones, specifically the recognition of word initial consonant clusters. For this recognition to take place, the speech stream that is the object of attention must be separated from its acoustic environment, which can comprise of other speech as well as traffic noise, animal voices (dogs barking, birdsong, etc.), music, non-speech sounds produced by humans (e.g., coughs, screams, yawns), and many more. To achieve this, listeners can exploit a signal's periodicity in addition to disparate frequency ranges within which the different audio signals occur, and group sound components with similar characteristics (e.g., several harmonics belonging to the same fundamental) and components with synchronous onsets and offsets together.[1] For example, listeners can use speakers' differing fundamental frequency (or pitch) ranges to divide the auditory signal into separate auditory

[1] For a detailed account of how listeners achieve this separation of auditory streams using *Auditory Scene Analyses*, see Bregman and McAdams, 1994.

streams—an effect generally known as the *cocktail party effect* (Arons, 1992). The product of the extraction of the attended auditory stream is a continuous speech stream in all its phonetic detail.

This detailed stream has to be both normalised and segmented (to create representations that can be matched with lexemes in the mental lexicon). During the normalisation process, the continuous stream is transformed into a sequence of discrete, abstract units (e.g., phones, see above). This is guided by listeners' knowledge of phonological rules (e.g., assimilation rules in a process called *compensation for coarticulation*), speaker characteristics, overall speaking rate, etc. The output of this step is a mental representation of the speech input that resembles a long sequence of discrete, abstract segments. In natural speech perception, this long chain then has to be segmented into lexeme-sized units in order to activate word candidates in the mental lexicon. Since the present study deals with the perception of monosyllabic nonce words, segmentation and lexical access will not be discussed here.

The following section gives a short overview of the factors involved in consonant and consonant cluster perception, and a summary of the findings of previous studies.

4.3. Perception of consonants and consonant clusters

At the lowest level of speech processing, the features that distinguish different phonemes, or rather their physical correlates, have to be extracted from the acoustic signal. To achieve this, listeners make use of so-called acoustic cues to correctly identify speech sounds. The relationship is not perfectly straightforward because phoneme contrasts are, by definition, abstract phonological (i.e., functional) distinctions, while acoustic cues take a concrete physical form. Nonetheless, several temporal, frequency-related, and intensity-related properties of the acoustic signal have been identified to serve as cues for phoneme identity.

Although auditory cues for phoneme identity are acoustically gradual, the result is the *categorical perception* of phonemes. This means that they are unambiguously perceived as belonging to one category or another, even when they are physically somewhere on a continuum with clear representatives of each category at either endpoint (Liberman et al., 1957; Repp, 1984).

The following paragraph gives a very brief summary of acoustic cues utilised for differentiating between consonants that are relevant in the present study. (Information taken from Reetz and Jongman, 2009, unless stated differently.) This background information is relevant for interpreting the experiment data and differentiating between the effects under consideration and lower-level acoustic effects. **Stop consonants** are easily identified by a period of silence (the closure) followed by a release burst and rapid formant transitions. Among them, different places of articulation are distinguished by the frequency of the release burst, as well as the formant transition pattern both into and out of the stop (the latter depends to a certain degree on the surrounding vowel, though). Voiced and voiceless stops are distinguished by the VOT value, closure duration, the intensity of the release burst, and the F0 and F1 frequencies of the following segment. **Fricatives** are characterised by frication noise and a relatively long duration. They are rather steady throughout their duration. The relative amplitude of the fricative and the location of the spectral peak serve to discriminate between different places of articulation, whereas the voicing distinction is determined by the presence vs. absence of low-frequency energy. Cues to **nasals** are weak formants (relative to vowels) with large bandwidths, a low-frequency resonance (the nasal formant), as well as anti-formants. The formant transitions into and out of the nasal are also characteristic but not necessary for the identification of nasals. Cues to nasal places of articulation can be found in both the so-called nasal murmur (characterised by a pattern of formants and anti-formants) and in transitions, but they are very complex. Therefore, identification of place is considered to be much harder for nasals than identification of nasal manner. The acoustic properties of **laterals** resemble those of vowels: there is a clear formant structure, which is weaker than in vowels but stronger than in nasals. Furthermore, there are anti-resonances. Formant transitions are longer than in stops. As Wright (2001; p. 254) notes, “not all cues [are] equally effective in conveying their information to the listener in all environments”. The strength of a cue may vary as a function of syllable position, phoneme environment, and presence vs. absence of noise (cf. Wright, 2001; and see below for a more detailed discussion). Moreover, the existence of a salient feature can partially obscure other contrasts. For example, Boersma (1998; p. 110) claims that the voicing contrast between [b] and [p] is perceived as stronger than the one between [v] and [f] because, in the latter, the frication noise distracts attention away from other contrasts.

Generally, no single cue seems to be necessary or sufficient for the identification of a phoneme. Rather, listeners build on several cues simultaneously to achieve identification with ease. If one cue is distorted or absent (due to abnormal pronunciation or masking by external noise), use of another cue leads to correct identification in many cases.

Moreover, cues to consonant identity are not only found in the segment itself but, due to coarticulation, can be carried by adjoining segments as well. Acoustic cues are therefore typically divided into *internal* and *external* cues. As the name suggests, internal cues are found in the phoneme itself; in contrast, external cues are located in the transition from and to the surrounding phonemes. The quality of external cues depends on whether the phonemes in question are good carriers of specific cues. For example, cues for place of articulation for stops lie as much in the transition to adjoining segments as in the stop itself. In fact, formant transition into a following vowel is considered to be the most reliable cue to stop consonant place of articulation (Henke et al., 2012). Generally, vowels are very good carriers of cues for preceding or following consonants, while most consonants are not. This means that perception of consonants in clusters can be very different from the perception of consonants surrounded by vowels. As a result, the findings of the studies reported above, most of which use CV or CVC stimuli, might not be transferable to consonant cluster perception.

There are a number of acoustic differences between singleton consonants and consonants in a cluster, all of which can affect their perception. For example, pre-vocalic voiceless stops are aspirated in German when they constitute a simple onset; when preceded by another consonant (as in /ʃp/), however, they are not aspirated and more closely resemble their voiced counterparts in physical form than singleton voiceless stops. [2] The transition to the following vowel preserves cues to place of articulation, however.

The situation is somewhat different for stops in C1 position. What makes them harder to distinguish is the lack of a clear formant transition to the vowel due to the intervening consonant in C2 position. As previously mentioned, formant transitions can serve as valuable cues for place of articulation in stops. When this transition is missing or obscured, the identification of place of articulation relies heavily on the frequency of the burst, which itself is easily masked by noise. Acoustically, stops in C1 position will therefore be more prone to masking by noise. Likewise, laterals and nasals depend heavily on transitions into vowels. Fricatives, on the other hand, are not as reliant on cues in the transition to a vowel and are thus less problematic in C1 position. Fricatives have very good internal cues to both place and manner (in the peak of stricture), so that they can be more easily identified, even

[2] As there are no contrasts in German between clusters with voiced vs. voiceless stops in C2 position, and this knowledge is acquired as part of the language-specific phonology, higher-level phonotactic and lexical knowledge should come into play here, thus delimiting confusability.

when followed by another consonant. Wright (2001; 265 f.) explains: "fricatives, and especially the sibilants, can still be recovered in the absence of a flanking vowel, whereas a stop will be more likely to be lost". From a perception perspective, sibilants are therefore much better candidates for C1 position in consonant clusters, which might very well be the reason why initial sibilant–stop clusters are relatively widespread cross-linguistically in spite of the fact that they violate the SSP (Henke et al., 2012). The consonant classes differ not only in whether they have internal acoustic cues for identification or whether those cues lie in the transitions to adjoining segments (preferably vowels), but also in their ability to carry cues regarding the identity of neighbouring consonants. As mentioned, vowels are ideally suited to carry information concerning neighbouring consonants. Laterals are also apt to carry such types of information (Pouplier, Marin, Hoole, et al., 2017), which makes consonant clusters like /pl/ less problematic than /ps/ with regard to recognition of the /p/. The overall perceptibility of a consonant's (internal and external) cues constitute its *cue robustness*.

When both a consonant's internal cues for its identity and its ability to carry cues for adjoining consonants are taken into consideration, consonant clusters can be classified as having a perceptually more or less favourable composition (cf. Pouplier, Marin, Hoole, et al., 2017; for an analogous classification of Russian onset clusters). Thus a classification scale for the perceptual ease of consonant clusters is proposed, ordered in terms of the following principles: C1 position should preferably be taken by a sibilant, and C2 position should preferably be taken by a stop, lateral, or nasal. Based on these considerations, the consonant clusters used in the present study can be classified as in Table 4.1.

Baroni (2014) further notes that a great acoustic difference between two consonants in an initial cluster—at least in the case of obstruent clusters—has a beneficial effect on its recognition: consonants that differ in both place and manner of articulation are recognized better than those that differ in only one feature. As there are no clusters that con-

optimal	medium	poor
/ʃt/	/pl/	/ts/
/ʃp/	/fl/	/ps/
/ʃm/	/kr/	/ks/
/ʃn/	/tr/	/tʃ/
/ʃl/		
/sk/		
/sp/		
/sl/		

Table 4.1.: Initial consonant clusters ranked according to their cue robustness

sist of two consonants of the same manner in the present study, this translates to a disadvantage for homorganic clusters (here: /ts/ and /sl/, plus /ʃt/ and /ʃn/ when alveolar and palato-alveolar places of articulation are equated).

Another acoustic difference between singleton consonants and consonants in clusters is that consonant clusters are articulated with relatively strong overlap in German, just like in English (as opposed to Russian, Pouplier, Marin, Hoole, et al., 2017), but the exact degree of overlap depends very much on the specific cluster (e.g., less overlap in /kn/ than in /kl/, cf. Hoole et al., 2009). As Wright (2004) notes, a certain degree of overlap might be beneficial for perception (because the cues are spread more widely), but if the overlap becomes too big, individual cues might be masked by others, which of course impedes identification.

As mentioned above, it has also been proposed that consonant clusters are perceived as holistic units rather than combinations of successive phonemes (Bond, 1971; Cutler et al., 1987; Treiman et al., 1982), which could be partly due to the strong temporal overlap.

In sum, consonants in clusters are acoustically distinct from singleton consonants. They are produced with relatively strong overlap. As many acoustic cues depend on transitions to a vowel, the perception of consonants in a cluster is very often more difficult than that of single consonants. This is also reflected in the fact that consonants in clusters are usually harder to detect in phoneme monitoring studies than singleton consonants (e.g., Treiman et al., 1982; although see Cutler et al., 1987, for relevant factors concerning this task). However, cue robustness depends very much on the individual phonemes involved and their order (i.e., position relative to the vowel). Therefore, different consonant clusters can be called "better" or "worse" based on their cue robustness. Furthermore, they might be perceived as holistic units rather than sequences of separate phonemes.

4.4. Models of speech perception

Before turning to the factors investigated in the present experiments and the experiments themselves, a brief digression to the theoretical underpinnings of phonotactic effects in speech perception is necessary. A number of speech perception models are able to account for sublexical frequency effects. In principle, all activation–competition models that feature a level for the sublexical unit under investigation should be appropriate. As has already been noted (Dell, 2000; Frisch, 1996), spreading-activation and connectionist approaches present a natural and cognitively realistic way to model gradient phonological phenomena in speech processing. In the following paragraph, two models that are particularly suitable for accounting for the kind of frequency effects studied here will be introduced. To the best of my knowledge, no speech perception model has been examined with respect to its applicability in accounting for sonority effects in speech perception and processing. Hence only brief and tentative suggestions will be made concerning their implementation. It is important to continually test these prominent models against empirical data of the kind reported in this thesis. The results of the perception experiment will therefore also be assessed with respect to their compatibility with these models.

4.4.1. The Neighborhood Activation Model and PARSYN

The Neighborhood Activation Model (Luce & Pisoni, 1998) is a mathematical model of spoken word recognition built around the idea that "words are recognized in the context of other words in memory". The basic assumption behind it is that a pure phoneme-by-phoneme matching process that leads to only one activated lexeme is unrealistic in light of the need for processing flexibility due to less than perfect input to the speech recognition system. Instead, the activation of a number of possible candidates and subsequent competition between them are postulated, as is characteristic of activation–competition models. Central to the Neighborhood Activation Model (hereafter NAM) are the structural relationships between items in the mental lexicon. They are assumed to be arranged in a multidimensional acoustic–phonetic space, whereby the dimensions correspond to phonetically relevant acoustic contrasts. Specifically, the most important factors for word recognition are the number, frequency, and phonological similarity[3] of the so-called neighbourhood, which consists of the activated candidates competing for recognition. When linguistic input is perceived, a number of *acoustic–phonetic patterns*, which need not necessarily correspond to real words, are activated in memory; their activation levels are a function of their similarity to the input. The acoustic–phonetic patterns corresponding to real words now activate a set of *word decision units* tuned to them. Upon activation, the word decision units begin monitoring the activation levels of the acoustic–phonetic pat-

[3] Phonological similarity, or confusability, was determined with the help of confusion matrices for position-specific phonemes from a CVC identification experiment in noise (Luce and Pisoni, 1998).

terns to which they are tuned, as well as the overall activity in the system and higher-level lexical information concerning the words they correspond to. Based on these three sources of information, the word decision units continuously compute probability values (so-called *decision values*) for the words they correspond to. This means that the word decision units integrate both bottom-up information which stems from acoustic–phonetic patterns and higher-level information into the calculation of decision values. However, priority is always given to the bottom-up information because acoustic–phonetic patterns activate the corresponding word decision levels in the first place. Lexical information, such as word frequency, can only act as a biasing factor by adjusting activation levels of acoustic–phonetic patterns according to the (log-transformed) frequency of the word they correspond to. The overall activity in the system equals the sum of the neighbour word probabilities, thus monitoring it reveals the amount of neighbourhood competition. When the decision value for a word decision unit reaches criterion, all information monitored by that unit is passed onto working memory, and the word is recognized.

Taken together, the recognition of a lexeme depends on the following factors: the intelligibility of the stimulus itself (as it activates the acoustic–phonetic patterns), its discriminability from other lexemes as indicated by neighbourhood density, and the frequencies of both the lexeme to be recognized and its neighbours. Crucially, frequency is not seen as an inherent characteristic of the activation levels of the acoustic–phonetic patterns but as a relative value that depends on the frequencies of all other activated units in the system, which can bias choice values towards a unit with a relatively higher frequency than its neighbours. Thus input corresponding to a HF lexeme with a sparse neighbourhood consisting of LF lexemes will be recognized better than input corresponding to a LF lexeme with a dense neighbourhood consisting of HF lexemes.

In the case of degraded stimulus input, it is assumed that no word decision unit will reach criterion; the decision for a word is then based on the values of the decision units once the processing of acoustic–phonetic information is complete. This assumption is relevant both for experiments with stimuli presented in noise and for naturalistic speech perception since it can account for the fact that facilitating effects of lexeme frequency (Cutler, 2012) and inhibitory effects of neighbourhood density (Vitevitch & Luce, 1998) and neighbourhood frequency are strongest under adverse listening conditions. It is here that the biasing effects of frequency are best brought to the fore because the activation of acoustic–phonetic patterns is least reliable.

NAM is a model capable of explaining a number of effects in speech perception related to frequency and word neighbourhoods, that is, the structural relationships between entries in the mental lexicon. It is therefore of great importance to usage-based linguistics. However, it lacks sublexical representations and instead features direct connections between acoustic–phonetic patterns and word decision units. This makes it inept in explaining a number of effects related to sublexical units, for example, perceptual learning effects and, crucial to this study, the simultaneous effects of sublexical frequencies and neighbourhoods. A few years after the postulation of NAM, however, Luce and colleagues developed a connectionist implementation of NAM, PARSYN (Luce, Goldinger, Auer, et al., 2000), which does feature sublexical units (namely position-specific allophones) and hence is capable of accounting for such effects.

In PARSYN, the external input (assumed to consist of position-specific allophonic units) activates units of allophone input as a function of their similarity to the input vectors. The allophone input layer has facilitative connections to a pattern layer, which is an exact duplicate of it, except for the fact that the units in the pattern layer have facilitative connections to pattern layer units in adjacent temporal positions, whereas there are no connections between units at the

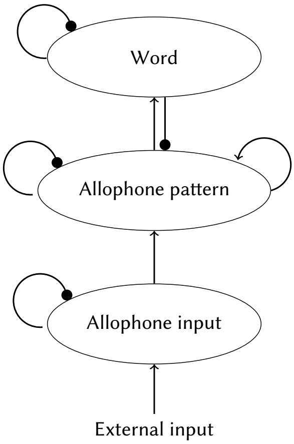


Figure 4.1.: Architecture of PARSYN (taken from Luce, Goldinger, Auer, et al., 2000; p. 619), arrows indicate facilitative connections, dots indicate inhibitory connections

input layer level. The weights of the lateral connections on the pattern level correspond to the log-frequency-weighted position-specific transitional probabilities.[4] This means that segments receive strong facilitation from segments they commonly co-occur with and less facilitation from segments they co-occur with less often. The resting levels of the nodes on the pattern level, on the other hand, correspond to the log-frequency-weighted (position-specific) probabilities of occurrence of those units. Taken together, the integration of these two kinds of probabilities—segmental and transitional—into the architecture of the model makes it very well-suited for explaining the effects of gradient phonotactics or sublexical frequencies on speech perception. The pattern layer is in turn connected to a word layer. Activated nodes of position-specific allophones send facilitative signals to the nodes of words the allophones appear in. The word units themselves, on the other hand, can inhibit each other, thereby manifesting competition. They also have inhibitory connections to the pattern units in order to reset the activation of the allophones once the relevant word has been

[4] PARSYN implements both forward and backward transitional probabilities, which makes it unique among models of spoken word recognition in its specificity with respect to predictions of phonotactic effects.

recognised. Word frequencies are taken into account by multiplying the word node activation values with log frequencies of the words they stand for. Figure 4.1 depicts the basic architecture of PARSYN. As can be seen in the figure, the architecture of PARSYN is quite similar to that of other interactive activation models of speech perception, such as TRACE (McClelland & Elman, 1986): there are inhibitory connections between within-level nodes and facilitating connections between interlevel nodes. There are, however, two exceptions to this rule concerning the nature of the connections in PARSYN: 1) As mentioned earlier, allophones in different temporal positions are connected by facilitative connections, and 2) the connections from the word level back to the pattern level are not facilitative but inhibitory.

When PARSYN is presented with an input word in the form of a sequence of position-specific allophones, the units at the allophone input level are activated. The strength of their activation expresses their perceptual similarity to the external input as estimated by confusion matrices. From there, activation spreads bottom-up to the other units in the model. The activation of a unit is calculated from its resting level activation and the positive or negative input the unit receives from all other units it is connected to by taking into consideration the respective connection weights, as well as its decay rate. The decay rate expresses how quickly the unit returns to its resting level; it is the same for input and pattern units, and twice as high for word units. The probability of PARSYN choosing a particular lexeme is calculated on the basis of its activation value in relation to all other activated word units' activation values.

In light of the aim of this dissertation—to examine and compare the influences of sublexical frequencies and sonority sequencing—and the sonority-related hypotheses laid out in Chapter 1, it is important to assess how sonority sequencing effects are or can be implemented in PARSYN. In the model as originally devised by the authors, sonority relations between phonemes are not implemented. In principle, though, PARSYN could be adapted to capture the notion of sonority effects instead of frequency effects. To achieve this, the connections between the units on the pattern level would have to be set according to their sonority values rather than their transitional probabilities. In the simplest case, connections to units in the following temporal position would be inhibitory if the transition to the phone which the units represents marks an SSP violation. For example, the unit representing the fricative /ʃ/ sends inhibitory input to all pattern units representing other (voiceless) fricatives or stops, thus penalising SSP-violating transitions, like /ʃt/. This implementation represents a binary distinction (violation/no violation) of the SSP. A more fine-grained implementation would involve weighting of the facilitative connections, which parallels the one implemented in PARSYN but for transitional probabilities. In this variant, pattern units send facilitative input to units in the following temporal position, with the weights of their connections corresponding to the rise in sonority. For example, the connection between the units /ʃ/ and /l/ would be stronger than the connection between the units /ʃ/ and /m/ because the sonority rise is bigger in /ʃl/ than in /ʃm/. No facilitative input would be sent from the unit representing /ʃ/ to the one representing /t/. Note that PARSYN was not designed to account for sonority effects, though. This is merely a suggestion for how its basic architecture can be used to account for speech perception effects beyond those of neighbourhood competition and frequency.

4.4.2. Adaptive Resonance Theory (ART)

Adaptive Resonance Theory (ART), developed by Stephen Grossberg (Grossberg, 1976), is a general cognitive theory of unsupervised learning. It models how the brain categorises, recognises, and predicts objects in a dynamically changing environment. Its main principle is a matching process between prior expectations and sensory input. Several ART variants have been developed to cover a diverse range of perceptual domains. The versions of ART that are relevant to speech perception are ARTPHONE (Grossberg et al., 1997) and its successors ARTWORD (Grossberg & Myers, 2000) and cARTWORD (conscious ARTWORD, Grossberg & Kazerounian, 2011). In the following section, a short overview of ARTPHONE will be given.

ARTPHONE models a neural network in which items (corresponding to bundles of acoustic features) in working memory (WM) are connected to so-called list chunks in short-term memory (STM). List chunks are representations of different-sized item groupings, such as phonemes, syllables, and words. In contrast to other models of speech perception, ARTPHONE makes no distinction between segmental, syllabic, and word levels; all of these units are represented by list chunks. It also differs from other models in that it is not the simple activation of a (word) unit that is the end state of perception, but a resonant state between an item in WM and a list chunk in STM. When input activates items in WM, these items send bottom-up activation signals via adaptive filters to list chunks that match their features. The list chunks then send top-down activation that corresponds to learned expectations (derived from long-term memory) of the pattern stored in WM back to associated item nodes. During this top-down matching process, expected items are selected for attentive processing and unexpected items are suppressed. Activated list chunks in STM that mismatch subsequent parts of the constantly unfolding signal lose their bottom-up support, while at the same time, they are inhibited by competitors; their resonant process is interrupted by a *mismatch reset*. As more and more of the speech signal arrives, the list chunks whose top-down signals (i.e., expectations) best match the incoming signal reinforce the items in working memory and, in turn, receive stronger bottom-up signals from them. It is this reciprocal activation that finally leads to a resonant state between the item and list chunk nodes, which is the aim of the perception process. In cases when no lexical chunks reach this resonant state, the "most predictive sublexical chunks" do (Luce, Goldinger,

& Vitevitch, 2000; p. 336). According to Grossberg (2003) "all conscious states in the brain are resonant states". Thus, when a resonant state between a list chunk and its associated items has developed, conscious recognition of that chunk occurs. A resonant wave travels across the network, which manifests the percept.

Like in most other connectionist models,[5] links between nodes of the same level (i.e., list chunks) are inhibitory in ARTPHONE. Therefore, the best-matching list chunks inhibit all other chunks. Moreover, ART features *masking*[6] of smaller list chunks by larger ones. This means that phoneme-sized list chunks are masked by syllable-sized chunks if the syllable contains the phoneme and matches the speech signal as a whole. This feature of the model can account for how listeners are able to perceive longer words that contain embedded words, as well as the divergent effects of phonotactic probability and lexical neighbourhoods for words and nonwords (cf. Vitevitch & Luce, 1999).

As an illustration, assume that the model, trained on the perception of German speech input, receives the input /plaːnt/ "plan(s)" (3rd person singular/2nd person plural). This would sequentially activate the items /p/, /l/, /aː/, /n/, and /t/ in WM. The /p/ item would activate, or *prime*, list chunks corresponding to *p*, *pa*, *pl*, *pr*, *pink*, *plus*, *Platz*, *Plan*, *plant*, and many more. The primed list chunks would send their activation back to /p/ and all other items that represent their components so that resonance begins to build up. As soon as /l/ is activated, the chunks that mismatch the updated input (*pa*, *pr*, *pink*, etc.) lose their bottom-up support, so that their resonances with /p/ are terminated by mismatch reset. The chunks *pl*, *plus*, *Platz*, *Plan*, and *plant*, on the other hand, are reinforced because the arriving speech signal matches prior expectations. However, the longer ones among them mask the

[5] with the exception of phonemes in different positions in PARSYN mentioned above

[6] Within the ART models, this term is used to refer to neuronal inhibition of chunks by larger chunks. It is not to be confused with the physical kind of masking by noise referred to later in this chapter.

shorter ones. As /a:/ is activated, only the chunks that match the input this far (in the example here, *Plan* and *plant*) are reinforced, while the others are inhibited. Since *plant* also inhibits *Plan* (by masking), resonances between the activated items and *plant* become strongest, which leads to a resonant state that constitutes the percept /pla:nt/. Figure 4.2 is a simplified schematic depiction of the processes. Please note that it does not show the state of the model at any particular point in time but rather several processes that take place consecutively. Here, resonances between items and several list chunks build up, but gradually the ones between the *plant* chunk and the input items become strongest, while some of the other resonances shown in Figure 4.2 are already terminated.

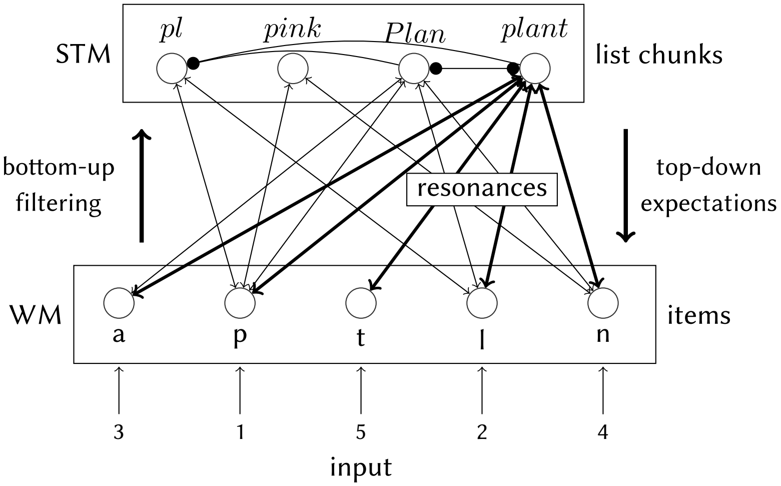


Figure 4.2.: Architecture of the ARTPHONE model (adapted from Grossberg et al., 1997; p. 484). Dots indicate inhibition, arrows indicate activation (double sided = resonance), thickness of arrow shafts indicates resonant strength.

It is the interaction between bottom-up priming and top-down expectations that enables ARTPHONE to model empirical findings, like the phoneme restoration effect (with the restored phoneme depending on the semantic context, Warren, 1984). What makes ART models particularly suited for modelling frequency effects is their integration of learning over a longer timespan as a source of expectations. It can account for how frequency distributions (in the present case, phonotactics) are acquired and why they affect speech recognition. Within the model, repeated encounters with a particular sound sequence leave traces and have consequences for future processing pathways:

> As incoming speech segments associated with words sequentially activate the[] item nodes, spatial patterns of activation evolve across the working memory. Repeated exposure to specific spatial patterns permits learning by the LTM traces in the adaptive pathways between the item nodes and the list nodes. (Grossberg et al., 1997; p. 488)

The adaptive pathways which send bottom-up activation from items to list chunks become tuned to the most frequent activation patterns over time, thus activating those list chunks more strongly than other patterns even though they might be equally consistent with the input. Specifically, it is resonant states between list chunks and items that trigger the learning of sensory and cognitive representations. This is a substantial part of the matching between top-down expectations and bottom-up activation: activation is enhanced for those chunks which, in the past, have proven to be important. As list chunks exist in different sizes, this principle provides an advantage for HF words, as well as HF syllables, biphones, and phonemes. In the example given here, the chunks *pl*, *plus*, *Plan*, etc. will be activated relatively strongly by the respective items because they constitute common activation patterns. If, on the other hand, the input to the system were /psalm/, the list chunks *ps*, *psa* and *Psalm* would receive weaker activation because they constitute less frequent activation patterns and, consequently, do not correspond to strong expectations. Likewise, the input /ts/ activates chunks like *ts*, *tsa*, *Zahn*, etc. more strongly than /psalm/ activates its associated list chunks. This is possible because the rate of presentation is not identical to the rate of recognition (i.e., the time the resonant wave takes to develop). Therefore, later segments can influence expectations for preceding segments (cf. Grossberg, 2003). Particularly when noise is present in the auditory signal, the activation of items in working memory might not be unambiguous and could lead to several activated items that share some of the features in the auditory signal. The resonant wave will be slower to develop, and later input with its activated items can influence the top-down expectations of previous items. With the activation of /s/, therefore, expectations concerning the previous item will be biased towards /t/ rather than /p/ because *ts* corresponds to the more common activation pattern.

According to ART models, the kind of learning effected by resonant processes continues throughout life. This prediction is consistent with empirical findings by Dell et al. (2000), who found that adult subjects are able to acquire artificial phonotactics during the course of an experiment. ART models are therefore well-suited to explain how language-specific phonotactics are acquired in the course of speech perception and how this knowledge is utilised to make subsequent speech perception processes more efficient. In contrast, it is unlikely that the principle of sonority or other (near-)universal rules of well-formedness could be integrated into an unsupervised learning model, like that of ART.

In an attempt to simulate the data that Vitevitch and Luce (1999) obtained in their experiments with human subjects in ARTPHONE, Pitt et al. (2007) found that the model could not only simulate the data in Vitevitch and Luce (1999) but also all other theoretically possible data outcomes. The prevalent pattern in their simulations, however, was the exact opposite of how the human subjects had behaved. Nevertheless, the lexicon they used for their simulations was very small (consisting of only four words, two of which were used as nonword input), and only two of ARTPHONE's nine parameters were used. It can therefore be theorised that this severe reduction led to the model's unexpected behaviour. All in all, ARTPHONE, with its interaction between bottom-up signals and top-down expectations, seems like a very promising approach to understand and model frequency effects—also in sublexical domains.

In this section, two connectionist models that seem particularly apt in accounting for sublexical frequency effects, PARSYN and ARTPHONE, have been introduced. In PARSYN, the main emphasis lies on neighbourhood competition, and phonotactic probabilities are encoded as connection weights. The idea behind ARTPHONE, on the other hand, is the integration of top-down expectations (which arise from previous experiences) and bottom-up activation. The acquisition of frequency distributions and subsequent utilisation of this knowledge are therefore inherent parts of the model's modus operandi. In the next chapter, a discussion—based on the results of the perception experiment—will address which characteristics of NAM and PARSYN accurately model the aspects of human sublexical speech perception investigated here and which ones are in need of revision.

5. Experiment 1: Native perception of German consonant clusters

5.1. Introduction

Native speech perception can be guided by top-down information. On a sublexical level, this information concerns the likelihood of different phoneme combinations. Whether this likelihood builds predominantly on language-specific distributions or is also informed to a considerable degree by universal phoneme sequencing regularities is not yet fully understood. The perception experiment reported here investigates factors influencing consonant cluster identification in noise, with the main focus on cluster frequency and sonority sequencing.

First, an overview of the relevant literature will be given in which the roles of acoustic factors, universal and language-specific phoneme ordering principles, as well as language-specific sublexical frequencies in native speech perception, are described.

5.2. Previous research

5.2.1. Consonant and consonant cluster perception

Most studies on phoneme perception have been concerned with acoustic correlations. Higher-level influences, such as that of phone frequency, have rarely been the object of interest, although Warner et al. (2005) and Benkí (2003) explicitly state that they did not find any phoneme/biphone frequency effects in their acoustically-oriented studies. In contrast, Moreno-Torres et al. (2017; p. 3089) note that, "when two consonants are spectrally comparable, phoneme frequency may explain the different patterns of consonant resistance". Although the phoneme as a unit of perception is problematic (cf. Foss & Swinney, 1973; Mitterer et al., 2013; Morais, 2021) and should probably be replaced by the phone, this is in principle a suggestion deserving of further research to bridge the gap between lower-level phone(me) perception studies on the one hand and studies investigating higher-level influences on the other hand. Studies examining the role of higher-level factors in speech perception, on the other hand, usually focus on the illegality of sequences (often in AX discrimination tasks, e.g., Tamási & Berent, 2015) or examine reaction time differences between sequences of different frequencies (Lentz & Kager, 2015). They do not test perceptibility differences between existing sequences under increased auditory uncertainty. The studies presented in this and the next chapter aim to bridge this gap by examining the identification of legal consonant clusters in noise and by taking into consideration both low-level acoustic factors in addition to two high-level factors, cluster frequency and sonority sequencing.

It is safe to assume that acoustic parameters play a crucial role in the identification of consonants in noise, but do language-structural factors, like cluster frequency and sonority sequencing, also influence consonant cluster perception in noise? Previous studies have shown that both high frequency of use of a cluster and its adherence to the SSP can facilitate cluster perception under certain circumstances. The following sections will give an overview of earlier findings concerning the influence of acoustic factors, frequency, and sonority on consonant and consonant cluster perception.

5.2.2. Acoustic factors

In addition to the numerous studies on categorical perception, most phoneme perception studies have been concerned with the interplay between different acoustic cues in phoneme perception or what makes some speech sounds particularly difficult to perceive, while others remain very robust even under adverse conditions. A lot of studies have used synthesized or synthetically manipulated speech sounds (like the classic studies by Ganong, 1980; Liberman et al., 1957; Massaro & Cohen, 1983), or noise-masked stimuli (e.g., Baroni, 2014; Cole & Iskarous, 2001); the presentation of natural, non-degraded stimuli in conjunction with reaction time measurements (with or without priming) has also been employed (Davidson & Shaw, 2012). Not surprisingly, most studies—especially the ones that employ a masking paradigm—have reported acoustic effects in perception (e.g., acoustic context: Cole & Iskarous, 2001; Massaro & Cohen, 1983; the role of the burst in stop place identification: Chang et al., 2001; acoustic landmarks: Silbert & Zadeh, 2015). A consistent result is that different consonants are affected by masking noise to variable degrees. There are several factors that determine a consonant's resistance to noise. Firstly, consonants with a higher inherent intensity—such as sibilants—are obviously more resistant to noise. In such cases, the frequency of the noise is also relevant: Moreno-Torres et al. (2017) found "energy above the masking noise [to be] the most important predictor of resistance". Secondly, aperiodic signals are more easily masked, which makes voiceless consonants more prone to masking.[1] Obviously, segments of short duration or with cues of short duration (e.g., release bursts of stops) are more easily missed than longer segments or segments whose identifying cues last longer (e.g., formants in vowels or formant-like structures in nasals, laterals, and glides; frication frequency in fricatives, nasal

[1] This might not be true in moderate noise conditions, however, in which transience may be more important (cf. Wright, 2001).

murmur in nasals), which makes their identification in noise even more difficult (cf. Wright, 2001). All of the above lead to an "asymmetrical degradation of information across manners of articulation" (Wright, 2001; p. 270). Stops and low-energy (i.e. non-sibilant) fricatives can be severely affected by noise, while laterals with their strong periodicity and especially sibilants with their inherently high intensity are more noise-resistant. Nasals group somewhere between these two extremes in most studies. The literature on the identifiability of different consonants in noise display great variability concerning the ordering of phonemes: this depends, inter alia, on the kind of noise used as well as signal-to-noise level (SNR), the language under investigation, the type of stimulus, and the use of closed vs. open sets of answers. A relatively robust finding is that noise has a more severe effect on the identification of place of articulation than on that of manner or voicing (e.g., Benkí, 2003; Warner et al., 2005; Woods et al., 2010).

Furthermore, an unfamiliar phonetic environment can impede consonant cluster perception since the acoustic cues for consonant identification are more difficult to recognise and parse in such cases (Davidson & Shaw, 2012). As described above, consonants in clusters differ in phonetic form from their singleton counterparts. If a consonant cluster is illegal in a given language, each of the consonants is situated in an unusual phonetic environment that leads to deviant acoustic cues. According to Davidson and Shaw (2012), one of the reasons why illegal clusters are so hard to perceive is that these cues cannot be properly interpreted by the listener. In this case, language-specific phonotactics exert an indirect influence on consonant cluster identification.

An additional acoustic factor relevant for the recognition of consonant clusters has been proposed by Baroni (2014): acoustic or perceptual salience (corresponding to what Henke et al. (2012) call *cue robustness*). On the basis of phonetic-acoustic factors, he identified a salience scale along which consonants can be located.[2] For example, fricatives are regarded as more salient than stops; among fricatives, sibilants are regarded as more salient than non-sibilants; among stops, dorsals as more salient than labials and both as more salient than coronals. His hypothesis is, “the more salient a consonant, the more easily it will be perceived correctly as the first member of an initial plateau cluster” (Baroni, 2014; p. 18); plateau clusters include both stop–fricative and fricative–stop clusters since he treats both kinds of consonants as obstruents. To meet his criterium of well-formedness in terms of salience, the first consonant of a cluster has to be more salient than the second one. His hypothesis was supported by the data for nasal and liquid plateau clusters but not for obstruent clusters, whose identification was better predicted by the legality of the cluster, the Net Auditory Distance between the two consonants (Dziubalska-Kołaczyk, 2009, 2014), and the phonetic context than the salience of the first consonant. Nonetheless, the hypothesis deserves further investigation, also with respect to obstruent clusters and clusters that do not constitute sonority plateaus.

5.2.3. Legality of phoneme sequences

There is a considerable body of research on the perception of consonant clusters—almost exclusively dedicated to the perception of phonotactically *illegal* clusters, that is, clusters that violate the phonotactics of the listeners’ L1. Converging results from psycholinguistic studies show that illegal tautosyllabic clusters are not perceived as accurately as legal ones (e.g. Berent et al., 2007; Davidson & Shaw, 2012; Newton, 1972). In many cases, an illegal consonant cluster is perceived to be a legal sequence. This phenomenon is known as a *perceptual assimilation* (Hallé & Best, 2007) or *perceptual illusion* (Berent et al., 2007).

[2] /s/ = 6, /ʃ/ = 5, /f/ = 4, /k/ = 3, /p/ = 2, /t/ = 1; /m/ = 1, /n/ = 0; /r/ = 1, /l/ = 0; numbers represent ranks rather than absolute values (Baroni, 2014; p. 23)

Listeners might, for example, break up an illegal cluster by perceiving an illusory vowel between the two consonants (e.g., *lbif* as *lebif*, cf. Berent et al., 2007; see also Pitt, 1998) or perceive the cluster as a phonetically close legal cluster (e.g., inital /tl/ as /kl/, Hallé et al., 1998; Massaro & Cohen, 1983; Moreton, 2002). Other kinds of perceptual assimilations, such as prothesis or consonant deletion, have occasionally been reported (Carlson et al., 2016; Davidson & Shaw, 2012). The kind of assimilation that occurs depends in part on the phonological properties of the cluster and the structure of the listener's L1.[3] In addition to this behavioural evidence, some neurolinguistic studies suggest that legal and illegal consonant clusters are processed differently and in partly different brain regions (e.g., Jacquemot et al., 2003; Rossi et al., 2011; Steinberg et al., 2016; Ulbrich et al., 2016; although see Raettig & Kotz, 2008).

The distinction between phonotactic legality and illegality is acquired early on in L1 acquisition: even at only nine months old, infants have been shown to be sensitive to legality differences in their native language (Jusczyk et al., 1993). It is, however, not immutable throughout life: later acquisition of a less restrictive phonotactic system gradually reduces the perceptual illusion effect in the more restrictive language, that means, the more proficient the listener is in the less restrictive language, the less susceptible he or she is to the perceptual illusion (Carlson et al., 2016). Another important factor for the relevance of legality seems to be stimulus quality: Massaro and Cohen (1983) found that the more ambiguous the stimulus was, the greater the phonotactic effect (see also McQueen & Pitt, 1996). At the same time, they report that "for a given goodness of match, a better match of the acoustic features is required for an inadmissible cluster than for

[3] As these results do not stem from naturalistic speech perception but from experimental paradigms (in many cases forced choice paradigms), the kind of illusion also depends very much on the exact experimental setup and the choice of alternatives offered by the researcher.

an admissible cluster" (Massaro and Cohen, 1983; p. 347). Thus their results suggest that bottom-up and top-down evidence can compensate for each other in the perception of consonant clusters. A study by Brown and Hildum (1956) showed that top-down influence is not restricted to simply knowing the phonotactic system of one's language, however. Rather, it generally reflects listeners' expectations, of which knowledge of the system is only one aspect. Brown and Hildum (1956) tested two groups of native English listeners in the perception of legal and illegal clusters; one group was informed about the mixture of legal and illegal onsets in the stimulus set, while the other expected only legal English syllables. Unsurprisingly, identification rates of illegal syllables in the informed group were more than four times as high as those in the uninformed group, which reflects the participants' expectations concerning the stimulus material. The authors interpret this result as an indication that listeners' expectations about structures are decisive for perception.

This stance is also taken in the present thesis: Speech perception is not solely determined by the acoustic input but also by the listener's (not necessarily conscious) expectations about it. The latter are informed to a considerable degree by the phonotactic structure of the language the listener is expecting to hear. This explains the aforementioned developmental trajectory quite well: as soon as an infant has become attuned to the phonotactic patterns of a specific language, he or she expects to hear them and is influenced by those expectations. After acquisition of a second, different phonotactic system, however, expectations are altered due to broadened experience. (For a model of how such learned expectations evolve, see Grossberg (1976, 2003) and Grossberg and Kazerounian (2011) and Section 4.4.2.)

5.2.4. Phonotactic probability

As has been summarised in Section 5.2.3, the effects of legality in consonant cluster perception are well-established. The picture is less clear with respect to more fine-grained phonotactic effects. In principle, phonotactics cannot only be categorical (differentiating legal vs. illegal structures), but can also be gradient, based on the frequency or probability of a given structure in the language. In German, the consonant sequence /lp/ is illegal when syllable-initial, while /pl/ and /ps/ are legal onsets; but /pl/ is the far more frequent of the two in the German language. This might have consequences for their processing in much the same way as categorical phonotactics. Gradient phonotactic effects in consonant cluster perception have been studied far less than categorical ones. The effects of illegal clusters as described above could be argued to simply be extreme cases of a basically gradient mechanism (cf. Cutler, 2012; p. 140). In that case, both categorical and gradient effects should be apparent but categorical ones, which represent the most extreme cases on a continuum of underlyingly gradient phonotactics, should be stronger.

One of the few studies that explicitly investigated effects of gradient phonotactic effects in the sense of *consonant cluster frequencies* is Pitt (1998). He carried out an identification experiment with synthetic liquid continua in consonant clusters of varying frequency. Here, no effects of cluster frequency beyond those of legality were found, i.e. the size of the phonotactic effect did not correlate with the frequency differences between the clusters. Pitt concludes that it is listeners' knowledge of permissible phoneme sequences rather than frequency that is relevant for consonant cluster perception. Likewise, Cohen et al. (1967), who found legality effects in both monolingual and bilingual populations, note in passing that a frequency-based account would leave many aspects of their data unexplained.

In contrast, Hay et al. (2004), who also investigated whether language-specific phonotactic effects are categorical or gradient, did find evidence of frequency effects on *heterosyllabic* (nasal–obstruent) clusters. In addition to a clear correlation between acceptability of the ratings of a cluster and the cluster's log frequency, they also report a frequency effect in reanalyses of the clusters: LF clusters were reanalysed more often than HF clusters and the direction of reanalysis was primarily from a less frequent to a more frequent cluster. They inferred that "[h]igh-frequency clusters attract responses, but only if they are acoustically similar to the speech signal." (Hay et al., 2004; p. 62). These results stand in contrast to Pitt (1998), who tested the perception of tautosyllabic and heterosyllabic clusters in separate experiments. The results converged and showed that there is no effect of cluster frequency over and above that of cluster legality (which was even more pronounced in the case of heterosyllabic clusters). Crucially, the two studies used different methods and different kinds of stimuli (acceptability ratings of minimally edited stimuli vs. identification of stimuli in synthetic consonant continua), as well as different cluster structures (nasal–obstruent vs. obstruent–liquid clusters). Any one of these differences might have led to the conflicting results. Choice of tasks and stimulus type and quality are known to significantly influence experiment results (cf., for example, Frauenfelder & Segui, 1989; Gerrits & Schouten, 2004; McQueen & Cutler, 2013; van Hessen & Schouten, 1999; Vitevitch, 2002).

So while Pitt (1998) and Cohen et al. (1967) claim that language-specific phonotactic effects in consonant cluster perception are categorical, Hay et al. (2004) propose that they are gradient, that is, based on cluster frequencies.

Finally, Lentz and Kager (2015) posit that categorical and gradient phonotactic effects in consonant cluster perception both exist but are fundamentally different in nature and mode of operation (see also Lentz, 2011; ch. 2). Probabilistic (i.e., gradient) phonotactics facilitate perception in frequent strings of phonemes, while categorical phonotactics might act as a filter to inhibit the recognition of illegal sequences. Their study focuses on an L2 learning scenario in which facilitative gradient phonotactics of the L2 can be acquired but through an L1 filter. This means that it is the filtered output (possibly including perceptual illusions when the input does not comply with L1 restrictions) that is fed into the gradient phonotactic learning process. In the scenario described by Lentz and Kager, the target of probabilistic phonotactic learning is *L2* phonotactics, while the categorical filter is determined by *L1* settings. Such a scenario can easily be transferred to a situation in which both categorical and gradient phonotactics are determined by the L1. Indeed, it follows from the logic of acquisition of gradient L2 phonotactics that knowledge of gradient L1 phonotactics should also exist and be acquired during L1 acquisition. In a purely L1-based scenario, categorical phonotactic knowledge would filter the input (which is useful for dealing with degraded or erroneous speech) and gradient knowledge would facilitate the processing of particularly frequent phoneme sequences.

Although the influence of gradient phonotactics on consonant cluster perception has not sparked much research interest thus far, the perceptual effect of phonotactic probability in general (i.e., calculated over the whole stimulus) has been investigated much more. Phonotactic probabilities are usually operationalised as either transitional probabilities (most commonly forward transitional probabilities, e.g., Pitt & McQueen, 1998; Yip, 2000) or position-specific segment frequency (e.g., Janse & Newman, 2012; van der Lugt, 2001). Vitevitch and Luce, who did the most extensive work on the matter (see below), integrate position-specific phoneme frequency and biphone frequency[4]—two highly correlated measures—into a single measure. In the following paragraph, the findings concerning all measures of phonotactic

[4] This comes closest to the (onset-specific) biphone frequencies used in the present study.

probability are collapsed for a better overview. There is no reason to believe that the precise measures used are important for the results reported here.

Most prominently, effects of phonotactic probability have been investigated in a number of studies by Paul Luce and Michael Vitevitch and colleagues (e.g., Luce & Large, 2001; Vitevitch & Luce, 1998, 1999; Vitevitch et al., 1999). They found clear evidence that phonotactic probability has a facilitative effect on word recognition and ascribe it to a sublexical level of processing. The distinction between sublexical and lexical levels of processing is central to their work as they show that there are opposing effects on both levels during speech processing; this easily obscures experimental results. Items of high phonotactic probability profit from facilitative effects on the sublexical level—the typical frequency effect that is manifested on many linguistic levels. However, high phonotactic probability is frequently correlated with high neighbourhood density (a high number of words that are phonologically similar to the item in question, see Section 4.4.1 below for an exact definition), which exerts an inhibitory effect due to competition on the lexical level. Hence a word with highly probable phoneme combinations enjoys facilitation from frequent phoneme sequences (sublexical level) and, at the same time, suffers from inhibition due to competing lexical representations (lexical level). However, the two effects do not behave in a linear fashion but interact in complex ways, depending on a number of factors, which include listener variables (e.g., subjective stimulus entropy: Luce & Large, 2001; hearing and attention switching abilities: Janse & Newman, 2012). Vitevitch and colleagues were able to demonstrate the separate effects by directing attention alternatively to the lexical or sublexical level (Vitevitch, 2003; Vitevitch & Luce, 1999; Vitevitch et al., 1999). This was done by using tasks that focus subjects' attention on one of the processing levels (e.g., lexical decision vs. same-different discrimination) or by using words vs. pseudowords[5]. Results consistently show that items of high phonotactic probability and neighbourhood density have a processing advantage when processing emphasis is on the sublexical level, but prove to be a disadvantage, when processing emphasis is on the lexical level. Both effects are therefore present simultaneously; and the visibility of either effect depends on whether the sublexical or the lexical level dominates processing in a given situation. In everyday speech perception, the lexical level usually dominates processing and effects of phonotactic probability are obscured by the dominant effects of lexical competition (cf. Vitevitch, 2003). The assumption of a sublexical and a lexical level of speech processing is adopted in the present study, but it is the sublexical level that is of primary interest given its facilitative effects of sublexical frequencies. To make these effects visible, it is crucial to control for effects of neighbourhood density and/or tap into sublexical processing by using appropriate tasks or pseudowords.

There is, to date, no consensus as to whether probabilistic phonotactic effects in speech perception truly reflect perceptual processes or whether they emanate from the representational level (cf. Auer Jr. and Luce, 2005, for a discussion of the literature). Neurolinguistic evidence suggests that these effects arise early in speech perception, that is, before definite contact to a lexical representation is made (e.g., Cheng et al., 2014; Pylkkänen et al., 2002), which supports a perceptual interpretation. Nevertheless, representational effects have also been found (Gathercole et al., 1999) and it seems very likely that phonotactic probability effects operate via both perceptual processes and representations. Moreover, McQueen and Pitt (1996) found effects of token frequency only for complex (CVCC) syllables, but not for simpler (CVC) syllables. They reason that distributional knowledge in the form of transitional probabilities might only be useful if perception is difficult to begin with.

[5] According to the authors, pseudowords do not initiate large-scale lexical competition because they do not make direct contact with any single lexical representation (Vitevitch & Luce, 1998; p. 328)

To sum up, there is only a very limited number of studies that investigate gradient phonotactic effects in consonant cluster perception and processing. The ones that do use a wide array of methods and reach different conclusions as to whether relative frequencies of legal consonant clusters affect their processing or not. Effects of probabilistic phonotactics in general, on the other hand, are relatively well-studied and have been shown to reliably influence speech processing once they are isolated from the opposing effects of lexical neighbourhoods.

5.2.5. Universal knowledge: Sonority

In addition to the acoustic and learning-related factors in the perception of consonant clusters, universal preferences for certain phoneme sequences—which also become apparent in well-formedness ratings—have been assumed to influence their auditory processing (e.g. Berent et al., 2007; Moreton, 2002; Tamási & Berent, 2015). Since such sequencing preferences are universal, they are reckoned to be based on language-independent structural knowledge of which sequences are "better" and which ones are "worse". Most of the time, the underlying principle is assumed to be sonority sequencing (see Section 3.3.2), which means that language users prefer syllables that are in line with the SSP and that this preference can affect speech processing. The behavioural distinctions that language users make between unattested phonological sequences on the basis of sonority are known as *sonority projections* (a term first used by Daland et al., 2011; and adopted by others).

For example, Moreton (2002) found that among various consonant clusters, all of which are illegal in the listeners' L1, the ones that are universally preferred are those that are easier to identify (see also Tamási & Berent, 2015; for a similar result). The authors of these studies conclude that this can be attributed to the universal phonological principle of sonority sequencing, which is the basis of perceptual difficulties regarding certain consonant clusters, rather than language-specific legality settings. Berent et al. (2007) note that this effect seems to be gradient: the more marked an onset cluster is, the more likely it is to be perceived as bisyllabic. Tamási and Berent (2015) specifically investigated sensitivity to sonority differences that are not distinguished in their participants' language: they tested English listeners on the sonority distance between stops and fricatives. As English phonotactic rules "do not systematically distinguish between the sonority levels" (Tamási and Berent, 2015; p. 362) of these two classes of phonemes,[6] any preferences that English language users might have concerning their combination is argued to be based on universal principles. In their analysis, they found no effect of language-specific statistical properties (similarity to the lexicon, neighbourhood, bigram frequency) but instead one of sonority distance between C1 and C2. Tamási and Berent (2015) do note that their results might be compatible with more sophisticated statistical explanations, such as the maximum entropy model (Hayes & Wilson, 2008), but point towards a "striking convergence between cross-linguistic preferences and the linguistic behavior of individual speakers [...] consistent with the possibility of universal grammatical restrictions on the phonological system".

Yet this cross-linguistic convergence is not necessarily as strong as the authors suggest. Studies by other researchers (e.g,. Daland et al., 2011; van de Vijver & Baer-Henney, 2012), on the other hand, show that the kind of sonority preferences observed by Berent et al. (2007) can, too, be derived from phonotactic distributions as long as they are described in terms of phonological features or natural classes as described above.

[6] As the authors acknowledge, /s/ is exempt from this observation since it is the only obstruent that can combine with another obstruent. This much discussed exception implies that /s/ is *less* sonorous than stops and therefore does not explain their results.

Not only can some intuitions that are usually ascribed to sonority be explained by generalised knowledge of language-specific distributions, but some findings lie in direct conflict with a sonority-based account. For example, Davidson and Shaw (2012) found significantly better recognition rates for fricative-initial clusters in syllable-initial position than for stop-initial clusters, while the opposite pattern was observed for final clusters by Bond (1971). From a sonority perspective, stop-initial clusters should be easier to perceive in syllable-initial position and fricative-initial clusters in syllable-final position. In addition, Davidson and Shaw's participants often confused test clusters with clusters that were *more* marked, in other words, the perceptual repairs did not improve the markedness of the clusters but deteriorated it. She concludes that a sonority account does not hold when the differences in sonority between the consonants in a cluster are much smaller than in the cases tested by Berent and colleagues. Instead, she attributes differences in perception accuracy to language-specific phonotactics.

However, there does seem to be an auditory-phonetic motivation for sonority sequencing which could explain its effect on speech perception: for accurate identification of a stop consonant, for example, a clear burst and a recognisable formant transition into the following segment are crucial (cf. Section 4.3). This is best ensured when the phoneme adjacent to the stop has a greater voicing amplitude and clear formant structure (Albright, 2007b). Wright (2001) demonstrates that more than half of acoustic cues to phoneme identity are dropped when a stop is followed by another stop (the most extreme case regarding sonority) as opposed to when it is followed by a vowel. These limitations explain the preference hierarchy of vowels, liquids, nasals, and obstruents following stops on purely perceptual grounds—an observation that famously led Henke et al. (2012) to ask: "Is the Sonority Sequencing Principle an epiphenomenon?"

Recent evidence suggests, however, that sonority as an acoustic prominence phenomenon and sonority sequencing as a structuring principle have separate facilitating effects on speech perception. In a study by Hamza et al. (2018), children with cochlear implants (CIs) were better at identifying nonce words with sonorous onsets (constituting a more gradual, i.e., dispreferred, sonority rise towards the nucleus but providing good perceptual prominence cues) than nonce words with less sonorant onsets (constituting better sonority sequencing but providing less perceptual prominence cues). Normal-hearing children, on the other hand, were able to utilise both sonority perceptual prominence and SSP cues in identification tasks and shift between the two. It seems, therefore, that restricting the role of sonority sequencing in speech perception to auditory cue prominence is premature.

In sum, the role of sonority in phonotactic knowledge and auditory speech processing is anything but conclusive. Nonetheless, there are strong indications that sonority is not as decisive for well-formedness judgements as often assumed and that many of its ascribed effects can be explained in terms of generalised learning of native structures. Crucially, most of the studies investigated metalinguistic judgements and the few ones that (successfully) tested effects of sonority-based phonotactics on speech perception did not control for alternative explanations, such as generalised probabilities. Given the plausible perceptual grounding of sonority sequencing, however, it is possible that the effect of sonority or the parameters correlated with it is stronger in automatic speech processing. Therefore, direct comparisons of the relative influences of language-specific (segment-based as well as generalised) and universal phonotactics are needed to make solid statements about the source of phonotactic influences in speech processing. The following section discusses the few studies that undertook such an endeavour.

5.2.6. Comparison of language-specific phonotactics and sonority

A study that compares the effects of sonority and language-specific phonotactic legality directly in a two-by-two design (albeit from a learning perspective) was conducted by Ulbrich et al. (2016). The behavioural and neurological (EEG) data from their CVCC nonce word learning experiment with speakers of German yield largely converging results that suggest that both legality and well-formedness regarding sonority sequencing are relevant to coda cluster processing.

In the behavioural data, the researchers found main effects for both sonority and German legality, as well as an interaction between the two: legal clusters were correctly identified more often if they adhered to the SSP, whereas sonority did not have an effect on illegal clusters. In contrast, legality did not facilitate recognition of sonority-violating clusters.

The ERP data were divided into two relevant time windows: 450-550 and 700-1050 ms post stimulus onset, respectively. Effects for sonority and legality in both time windows were only significant in expanded statistical models that included acoustic variables.[7] In reduced models without acoustic parameters, the main effect of legality disappeared altogether, while the main effect of sonority only occurred during the first time window. It surfaced as a negativity effect for SSP-violating clusters, which is interpreted as a pre-lexical form-based analysis of the clusters.

The interaction between legality and sonority was significant in both time windows (in all models). The interaction within the first window indicated that “[c]onflicting information of two competing factors influencing the processing of words may increase processing costs. [...]

[7] The expanded models actually provided the best fit of the data but were not analysed further because the large number of parameters made them difficult to interpret.

The processing of marked structures may be more difficult when they exist and lead to deeper processing" (Ulbrich et al., 2016; p. 674). During the second time window, on the other hand, the effect (this time a positivity effect) was strongest for clusters that violate both sonority sequencing and phonotactics. Consequently, legality and sonority interact in diverse ways during different ERP components (with the first time window interpreted as an N400 effect and the second as a late positive component, LPC). Ulbrich and collaborators conclude that both sonority and legality play a significant role for processing word-like structures but that sonority is the more important of the two. They reason that language users might profit from language-specific phonotactics when exposed to structures that conform to universal phonotactic constraints, but not to those that violate them. However, they emphasise that both principles are not mutually exclusive (as is often presumed in the literature).

A parallel study with Polish speakers was conducted by Wiese et al. (2017). On the whole, the results resemble those of Ulbrich et al. (2016), although there were no main effects for either sonority or legality in the behavioural data, nor in the EEG data for the first experimental session prior to learning. During the second session, after the learning period, both sonority and legality showed effects during processing. One of the most interesting findings from the study is that these effects emerge at different points in time: while sonority violations lead to early increased negativity (during the first time window), the effects of legality of a cluster occur later (positivity during the second time window). Just as in Ulbrich et al.'s (2016) study, even legal clusters lead to different brain responses depending on their sonority status. The authors deduce from these results that processing sonority, which is based on properties of the acoustic signal, might act as a filter for "fast and relatively effortless perception of relevant clusters" (Wiese et al., 2017; p. 12). Processing legality, on the other hand, involves access to a phonological lexicon and therefore shows later effects. Since only sonority exhibits a main effect, while effects of legality are only revealed in interactions, the authors conclude that sonority makes a greater and more direct contribution to speech processing. Given that the most interesting effects only occurred during the second EEG session, however, one should be cautious in interpreting these results independent of the learning scenario that was the focus of interest (even though the authors do not seem to impose this restriction on their own interpretation).

Wiese et al. also note—since the study was conducted with speakers of Polish, who are exposed to a relatively large number of sonority-violating clusters in their native language—that sonority is a principle that exerts an influence even in the absence of evidence from the input. However, like Ulbrich and colleagues, they do acknowledge that both principles—sonority and legality—influence how the adult brain processes and learns language structures.

In contrast, a similar study in which Russian and Chinese learners of German were exposed to CVCC syllables found that Russian learners were not significantly influenced by either sonority or legality (in German), while the Chinese learners were influenced by sonority and advanced Chinese learners by both legality and sonority (Ulbrich & Wiese, 2018). This indicates indirect influence from the phonotactic system of the native language: the Russian group, which is used to many different consonant clusters (some of them sonority-violating), had less trouble processing the clusters than the Chinese learners, whose native language does not allow any coda clusters at all. This shows how sonority sequencing, legality (in L2) and the relative restrictiveness of the L1 interact during processing.

Hence the studies comparing the effects of (il-)legality and sonority directly show that both can influence speech processing, although sonority seems to exert a stronger and more direct influence, at least with respect to learning. The effect of sonority seems to partly depend on the listeners' L1 phonotactic system. However, all of the studies discussed here interpret both universal and language-specific phonotactics as binary: a cluster is considered either legal or illegal in a given language and either well-formed or ill-formed on the basis of sonority. In reality, both sonority-based and language-specific distributional phonotactic knowledge can, in principle, be categorical or gradient. Regarding sonority, one can draw a clear distinction between sequences that conform to the SSP and those that violate it. On the other hand, phoneme sequences can be closer to or further from the ideal sonority distance—thus exhibiting a gradient metric. For distributional phonotactics, gradience is even more evident as it can be captured (an alternative to the dichotomous distinction made by Ulbrich and colleagues) by frequencies.

To the best of my knowledge, there is very limited research comparing gradient effects of language-specific phonotactics and sonority on consonant cluster perception. Some insight can be gained from Albright (2007b), who tested the effects of both generalised language-specific probabilities and a prior sonority bias on phonotactic acceptability ratings. He found that both contribute significantly to the model's prediction of human ratings and concludes that "the best available model for the data is one that incorporates both inductively learned constraints (reflecting statistical properties of English) and also prior constraints (reflecting a universal preference for stops to be followed by more sonorous segments)" (Albright, 2007b; p. 3). The existence of separate and complementary effects of language-specific and universal phonotactics on acceptability ratings[8] should also be tested with respect to (facilitation in) speech processing. This is exactly what the experiment described below is aimed at.

[8] Note, however, that Albright used the consonant clusters from Scholes's (1966) seminal study, a wide array of clusters comprising of both attested and unattested clusters over the whole sonority range. It is not unlikely that the two predictors explained different areas in the space of consonant clusters.

5.3. Motivation of the present study and hypotheses

To sum up, psycholinguistic research on phonotactics has unveiled effects of both language-specific phonotactics—categorical as well as gradient—and sonority-based preferences on speech perception and metalinguistic judgements, the latter of which may be grounded in phonetic principles (e.g., that stop-initial consonant clusters require a more sonorous phoneme in C2 position for a stop's acoustic cues to be retained). For learning, a categorical sonority distinction seems to be a more powerful influencing factor than categorical language-specific phonotactics. When taking gradient phonotactics into consideration, they can account for a larger part of well-formedness intuitions than sonority-based predictions—especially when they include generalised frequencies in addition to segment-specific ones. Nonetheless, even generalised language-specific phonotactics cannot fully explain human phonotactic ratings. If they are complemented with a prior bias for phonetically motivated sonority principles, however, human ratings can be captured fairly accurately. For accuracy in native auditory perception, no similar comparison involving gradient phonotactics has been made.

To test if a combination of the aforementioned phoneme sequencing metrics is equally influential when it comes to automatisation in speech processing, an auditory identification experiment with onset clusters of different frequencies as well as different sonority categories was conducted. In order to avoid ceiling effects, stimuli had to be masked. Since white noise has differential effects on various frequency bands, multi-talker babble was chosen as a masker. To rule out effects of lexical frequency and familiarity, only pseudowords were used.

In keeping with the general hypotheses of this dissertation, the following hypotheses are formulated for the native perception experiment:

1. High cluster frequency will facilitate perception. This means error rates should be lower for HF clusters than for LF clusters.

2. Since high-probability outcomes are favoured in situations of uncertainty, errors should result in the percept of a HF cluster. More specifically, they should result in the percept of a cluster with a higher frequency than that of the target more often than the percept of a cluster with lower frequency.

3. Consonant clusters that conform to the SSP should be perceived more accurately than clusters that violate it. Error rates should therefore be higher for clusters violating the SSP.

4. Moreover, errors are hypothesised to improve the sonority profile of a cluster rather than deteriorate it. Essentially, this means that SSP-violating clusters should tend to be perceived as SSP-conforming clusters and, generally, the sonority distance between C1 and C2 should be increased rather than decreased.

5. It is hypothesised that linguistic experience is more relevant to perception than sonority sequencing. The effect of frequency should therefore be stronger than that of SSP violation and, in cases where the two make diverging predictions, frequency is thought to be the better predictor. For example, error rates for /tʃ/ (LF, SSP-conforming) should be higher than for /ʃt/ (HF, SSP violation).

6. Frequency and sonority sequencing might interact, leading to stronger effects of sonority sequencing for LF clusters, while the overlearnedness of HF clusters would reduce the sonority effect for them.

5.4. Methods

5.4.1. Participants

35 native speakers of German (22 female, mean age: 24.06, SD = 4.10), mostly students at the University of Freiburg, were tested and paid for their participation. None of them reported any hearing impairment.

In order to obtain listeners with a comparable language background, the call for participation explicitly ruled out participants that speak a dialect, as well as participants from the south of Germany (who have at least passive experience with the cluster /ks/ that goes beyond that of Standard German speakers).

5.4.2. Materials

Stimuli

The 16 test clusters presented in Section 3.5.2 were used. For convenience, they are listed here along with their log CELEX type frequencies (Table 5.1).

cluster	frequency	cluster	frequency
/ts/	3.25	/pl/	2.23
/ʃt/	3.16	/ʃn/	2.18
/ʃp/	2.91	/sk/	1.94
/tr/	2.88	/ps/	1.54
/kr/	2.61	/sl/	1.36
/ʃl/	2.54	/tʃ/	1.11
/fl/	2.40	/ks/	0.95
/ʃm/	2.25	/sp/	0.85

Table 5.1.: CELEX type frequencies (log-transformed) of the 16 test clusters

160 monosyllabic pseudowords beginning with the test clusters were created. Pseudowords rather than real words were used in order to

avoid lexical effects; for the same reason, they were lexically opaque (cf. Raettig & Kotz, 2008). At the same time, the stimuli were created to sound as natural as possible in German. Fifteen different nucleus vowels (/a, aː, eː, ɛ, iː, ɪ, oː, ɔ, uː, ʊ, øː, œ, yː, ʏ, aɪ, aʊ, ɔʏ/) and 13 different coda consonants (/p, t, k, f, s, ʃ, ç, x, m, n, ŋ, l, r[9]/) were used and distributed as evenly as possible over the stimuli in general, over onset clusters, and over experimental blocks. (Diphthongs and umlauts were used less than base vowels: Diphthongs are used in 31 stimuli, umlauts in 22, other vowels in 197.) Altogether, each onset cluster was combined with each base vowel (either as a long or as a short vowel or with both variants) and with an umlaut or diphthong to yield 10 different onset–nucleus combinations. Coda consonants were then added to make pronounceable syllables that are not too closely/immediately associated with any real German word. All test stimuli had the form CCVC (with long or short vowel, e.g., /ʃteːm/, /flœp/) or CCVVC (e.g. /ksaɪn/). In addition to the 160 test stimuli, 100 filler stimuli (also monosyllabic pseudowords) were created, 73 of which had simple onsets. The remaining 27 filler items started with consonant clusters that were not part of the test set (e.g./kluːf/, /ʃvøːt/). A full list of stimuli can be found in Appendix A.

The stimuli were spoken by a trained female native speaker of German with standard pronunciation and recorded in a sound-attenuated booth with an AKG C2000B microphone in Adobe Audition. They were recorded in stereo channel with a sampling rate of 44,100 Hz. Each stimulus item was spoken at least three times and the best token of each item was selected.[10] After token selection, all stimuli were RMS-normalised to 65 dB Sound Pressure Level (SPL) in Praat (Boersma

[9] In natural conversation, this is mostly pronounced as tiefschwa /ɐ/ in coda position. In the stimuli, it was articulated as a consonant (velar fricative) in order to keep the syllable structure consistent.

[10] One stimulus, /ʃpɛf/, was created by splicing two naturally recorded stimuli (/ʃpɛl/ and /spɛf/) because it had been mispronounced in all recordings. Both original recordings were cut at a zero crossing of the amplitude wave: the spliced stimulus

& Weenink, 2018) using a script from the Northwestern University Linguistics Department (http://groups.linguistics.northwestern.edu/speech_comm_group/documents/LTAS/normalize_audio.praat). As this normalisation relies to a large part on the vowel of the stimulus, the onsets were only approximately equal in intensity. There was no feasible way to normalise only the onsets of the stimuli. However, as different types of consonants differ inherently in their intensity, this can be seen as increasing ecological validity. The intensity of the stimulus onsets varied between 52.2 and 69.0 dB (SPL). Onset duration ranged from 81 ms to 333 ms. Both intensity and duration were significantly different for some of the clusters (see Figure A.1 in the Appendix), as is to be expected in light of the natural variation between consonant classes.

Multi-talker babble

Twenty recordings of German audio books (10 male and 10 female readers) were taken from a website for public domain audio books (https://librivox.org), which served as the source for the multi-talker babble. The number of babble talkers was 20 because during internal pretesting, 6- to 8-talker babble, which are the numbers of babble talkers most widely used (e.g., Cutler et al., 2008; Felty et al., 2013; Heinrich et al., 2010; Treiman et al., 1982), turned out to be too variable in intensity and pitch over short stretches of time. This resulted in a strongly varying masking effect and a high rate of informational masking with inadvertent recognisability of words in the babble. The optimal number of talkers was then determined by auditory checking of different variants until the babble sounded uniform and almost no individual words could be recognised. The 20 recordings were chosen based on audio quality, clarity of pronunciations and steady intonation, and were pro-

sounded perfectly natural and did not deviate in any noticeable way from the other stimuli.

cessed as follows: 1) removal of silences above 0.15 s in Praat, 2) RMS normalisation to 65 dB SPL, 3) mixing of individual files in Audacity software (Version 2.1.1, http://audacityteam.org), 4) trimming to length of the shortest source file (to ensure that all talkers were present at a given time) and exclusion of initial 15 s, which contained general announcements and the title, 5) cutting into pieces of 2 seconds, 6) final normalisation of babble pieces to 64 dB SPL, and 7) auditory checking of all babble pieces and elimination of pieces with noticeable intensity peaks.

Normalising the stimuli to 65 dB SPL and the multi-talker babble to 64 dB SPL yielded a signal-to-noise ratio (SNR) of +1 dB SPL.[11] The SNR is based on observations from a pre-test with four participants, none of which participated in the final experiment. The stimuli and procedure were almost exactly as described below for the final experiment, with three exceptions: Firstly, the SNR was varied between +3 dB SPL (*n* = 2) and +1 dB SPL (*n* = 2). Secondly, the experiment version for two of the participants included the test clusters /sp/, /ks/, /tr/, /sl/ and /t͡s/ in practice trials, while in the final experiment, none of these test clusters were used in practice trials. And thirdly, all pre-test participants were instructed to spell /sp/ as <ßp> (<ß> being a common German letter to denote /s/, which is, however, never used in syllable-initial position). This was changed to <s-p> for the final experiment (see below).

During the pre-test, the SNR of +1 dB SPL led to error rates of 31.9% and 43.8%, which provides enough data for analysis while not simultaneously frustrating participants. The SNR of +3 dB SPL, in contrast, yielded error rates of 14.4% and 28.8% in the pre-test, which could lead to ceiling effects.

[11] It is worth noting that the SNR was determined based on the mean intensity of the whole stimulus. Stimulus syllables were normalized to a mean intensity of 65 dB. Hence, the actual difference in intensity between signal and noise at the onset of the stimulus (which is the sole determiner of accuracy) varies. Mean intensity of the stimulus onsets is 61.22 dB SPL (leading to a mean SNR of -2.78 dB SPL at syllable onsets), with a standard deviation of 3.48.

5.4.3. Design and procedure

Before the experiment, all participants completed a questionnaire on their personal data, language background and experience with German dialects, as well as potential hearing impairment (see Appendix A). For the experiment, subjects were seated one at a time in a sound-attenuated booth and equipped with headphones.

The task was open-set recognition, which means that it involved free transcription of the stimuli. The 100 filler items described above were used to ensure that participants did not recognise the task as closed-set recognition and only use answers from the set of test clusters.

The experiment was run in OpenSesame 3.1.9 (Mathôt et al., 2012) on a MacBook Pro. It was divided into five blocks of 52 stimuli each. After each block, a screen informed participants that they had completed the *n*th block, and instructed them to take a short break and then press the enter key when they were ready for the next block. The order of the blocks was counterbalanced across participants using a Latin Square design and the order of the stimuli within each block was pseudo-randomised by the experiment software. Pseudo-randomisation included the following constraints: 1) No test cluster can appear in two consecutive trials. 2) Not more than three test items or three filler items can appear in succession.

Subjects were instructed orally to listen to the nonsense syllables in babble and type what they heard when they saw the prompt on the screen, make corrections where needed and press the enter key when they were ready for the next stimulus. They were told that the task demanded a lot of concentration and they should only press the enter key when they were fully focused. They were instructed to listen very closely and write down exactly what they had heard. They were asked to base their transcriptions on German spelling conventions but additionally received a sheet showing the desired spelling system by means of examples (e.g. notation of long vs. short vowels and <s-p> to denote /sp/ since <sp> is conventionally used to denote /ʃp/ in German orthography). Participants studied this sheet and were free to use it during the experiment as needed. They were not informed that the focus of interest was on the syllable onsets, so instructions concerning spelling included all syllable parts. During the introduction, participants could ask questions for clarification at any time. After the oral instructions, the crucial points were repeated in written form on the computer screen. Here, audio examples of possible stimuli (without noise) were given along with specification of the desired spelling. To increase motivation for exact transcriptions, the person with the least mistakes was promised a gift card as a reward in addition to the monetary compensation that all subjects received for participation.

Before the syllable identification task started, a short hearing screening was performed. Fifteen pure tones at five different frequencies (500, 1000, 2000, 4000, and 8000 Hz) were generated in OpenSesame and played over headphones at random intervals at the lowest amplitude possible (approximately 25 dB SPL). Subjects were instructed to press the space key as soon as they heard a tone. There was a timeout after 5 seconds. This test does not meet clinical criteria but was conducted to eliminate an influence of acute or permanent hearing impairment on consonant identification. Two participants missed one tone each (500 and 1000 Hz, respectively). Since their error rates in the experiment proper were below the overall mean error rate and none of the other participants missed a tone, it can be concluded that general perceptual deficits did not influence the experiment results. Declarations in the self-reports support this conclusion.

Ten practice trials were then given to familiarise participants with the task and the desired spelling of the nonce syllables. During the practice trials, participants received feedback on the computer screen to indicate that that their answer was correct or to show the correct answer spelled out. Participants were encouraged to set the volume at a comfortable listening level during the practice trials and leave it constant throughout the experiment proper. One participant repeated the practice trials at his own request.

During each trial, the stimulus was played only once and could not be repeated. Multi-talker babble started 415 ms before the onset of the stimulus and continued after the offset of the stimulus for a total of 2 seconds. The stimuli were randomly assigned to the 260 babble segments for every participant anew to minimise the effect a given babble segment may have on a stimulus. Immediately after the offset of the audio, the question *Was hast Du gehört?* "What did you hear?" and an input field appeared on the screen, indicating that the participant could start typing. Participants saw their typed responses on the screen and could correct them if desired before pressing the enter key. When they pressed the enter key, the next trial was initialised. That means the experiment was fully self-paced. The whole experiment lasted about 45 minutes.

5.4.4. Analysis

The dependent variable in the experiment was the error in the onset cluster (binary variable), which means that only the part of the collected written answers that preceded the first vowel was rated for correctness. Onsets were transliterated into phonetic writing (SAMPA) and compared to target onsets to obtain error counts. Both an error (deletion or substitution) in only one of the onset consonants and an error concerning both consonants or a consonant addition were counted equally as one mistake. Transcription of a voiced consonant instead of its voiceless target counterpart (e.g., <bs> instead of <ps>) was not counted as a mistake in obstruent-obstruent clusters because the stop in such a cluster is phonetically closer to a voiced stop (cf. Section 4.3).

It turned out that /sp/ was spelled <sp> (which was meant to denote /ʃp/, see above) instead of <s-p> much more often than hearing errors could plausibly account for. This would appear to be attributable to difficulties in adhering to the spelling system set up for the experiment. Due to these assumed "spelling mistakes" the number of errors for /sp/ was disproportionately high. For this reason, all cases in which the stimulus cluster /sp/ was transcribed as <sp> had to be excluded from the analysis (thereby probably also eliminating some cases of true hearing errors). It was decided to exclude only the cases in which the cluster was spelled as <sp> and not the cluster as a whole because this cluster is very valuable to the experimental setup. It is the only sonority-violating cluster with a very low frequency and therefore represents an important cell in the experiment design. If all cases of /sp/ error had been submitted to the statistical analysis, on the other hand, the observed effects would have been unjustly exaggerated. Therefore, by only excluding the cases in which /sp/ was transcribed as <sp>, this was deemed to be the best compromise, leaving true hearing errors for analysis. Moreover, two cases had to be excluded from the analysis because they contained empty responses. After this procedure, 5452 out of the original 5600 observations were left in the data set (exclusion rate: 2.64%).

Logistic regression

The data were analysed in a logistic regression mixed-effects model fit by maximum likelihood (Laplace Approximation) with the lme4 package (Bates et al., 2015) in R (R Core Team, 2016). Perception error in the onset served as the dependent variable. Model fitting proceeded in stepwise forward fashion with the minimal model containing only log cluster frequency and sonority violation as fixed effects.

Log cluster frequency is based on CELEX type frequencies and was calculated as described in Section 3.5.2. Sonority violation was used as a categorical variable with the levels "no violation" and "violation". To this minimal model the following predictors were added stepwise and model fit compared using the *anova* function of the lme4 package:

summed frequency of neighbour clusters, well-formedness in terms of salience, generalised cluster frequency based on natural classes, onset duration, onset intensity, and NAD difference between CC and CV transitions (based on Dziubalska-Kołaczyk, 2014).

To calculate neighbourhood frequencies, all legal German consonant clusters neighbouring a cluster in onset position were compiled and their log type frequencies summed. A cluster was defined as a neighbour if it deviated from the cluster under consideration by one phonological feature (place of articulation, manner of articulation, or voicing) in one of the consonants.[12] This was the closest sublexical equivalent of the frequency-weighted neighbourhood-density value used by Luce and Large (2001) and Vitevitch and Luce (1998, 1999; see Section 5.2.4), although their definition of a neighbour is based on addition, deletion, or substitution of one *phoneme*. Conversely, here a neighbour is defined as a change in one *feature*. Changing a whole phoneme was considered to be too coarse-grained in the case of sublexical units that are just two phonemes long.

Salience-based well-formedness (based on Baroni, 2014) was operationalised as a factor with the levels *well-formed* (meaning C1 is more salient than C2 according to Baroni's scale) and *ill-formed* (C2 is more salient or both consonants have the same value; pertaining to /ts, tʃ, ps, ks ,kr/).

To calculate generalised cluster frequencies, all German onset clusters were categorised according to the manner of articulation of each individual consonant. Fricatives were subdivided into sibilants and non-sibilants (i.e., /f/), because the former are known to display perceptually and structurally distinct behaviour (Baroni, 2014; Henke et

[12] The transitions from /ʃm/ to /ʃp/ and from /ʃn/ to /ʃt/ were counted as one-feature changes because the stops in C2 position are phonetically closer to voiced stops, thus there is only a change in manner of articulation. Furthermore, the transition from /tʃ/ to /d͡ʒ/ was also counted as a change in just one feature because a change of voicing in one of the consonants necessitates a change of voicing in the other as well.

al., 2012). The log frequencies of a category's members (based on WebCELEX and derived as described in Section 3.5.2) were then summed to obtain the generalised frequency. Table 5.2 lists cluster categories along with their members and generalised frequencies.

cluster generalisation	member clusters	generalised frequency
stop–liquid	pr, **tr**, **kr**, br, dr, gr, **pl**, kl, bl, gl	21.571554
stop–sibilant	**ts**, **tʃ**, **ps**, **ks**	6.799551
sibilant–stop	**ʃp**, **ʃt**, **sp**, st, **sk**	8.803511
sibilant–nasal	**ʃm**, **ʃn**	3.837715
sibilant–liquid	**ʃl**, ʃr, **sl**	5.556418
fricative–liquid	fr, **fl**	4.318606

Table 5.2.: Generalised consonant cluster frequencies; test clusters in bold

Onset duration was measured individually for each stimulus in Praat. For that purpose, onsets were selected based on spectrograms and acoustic inspection. Onset intensity (in dB SPL) was also measured in Praat, based on the same selections as used to measure duration.

NAD difference was calculated with the online NAD calculator (http://wa.amu.edu.pl/nadcalc/ Dziubalska-Kołaczyk et al., n.d.) specifically for German by subtracting NAD(C2–V) from NAD(C1–C2). The vowel was left unspecified and the option to include sonority in the calculation was chosen.

It seemed plausible that frequency and sonority could interact, which might result, for example, in stronger sonority effects for LF than for HF clusters. As a result, an interaction term between the two predictors was also included.

All numerical variables (cluster frequency, generalised cluster frequency, neighbour cluster frequency, onset duration, and onset intensity) were centred before being entered into the model. For all categorical variables, sum coding was used.

The best fitting model included log cluster frequency, generalised cluster frequency based on natural classes, summed frequency of neighbour clusters and intensity of the onset as numerical fixed effects and SSP violation as well as cluster well-formedness in terms of salience as categorical fixed effects. The parameters NAD difference and onset duration did not improve the model, nor did including an interaction between frequency and SSP violation. The final model also included random intercepts for subject, stimulus and onset cluster (the latter two nested), and random slopes for log cluster frequency, SSP violation, neighbourhood frequency, salience-based well-formedness, and onset intensity by subject. Including random slopes for the remaining fixed effects by subject did not improve the model. No random slopes by stimulus or onset cluster were included in the model as the fixed effects did not vary within a stimulus or onset cluster.

To see whether the sonority effect is better captured by a finer measure of sonority sequencing, an additional model was run in which the binary sonority variable was replaced by a categorical variable with four levels (-1, 1, 2, 3) to represent the sonority distance between C1 and C2 in a separate model. Table 5.3 shows the sonority distance values for the individual onset clusters.

onset clusters	sonority distance
ʃp, ʃt, sp, sk	-1
ʃm, ʃn, ps, ts, ks, tʃ	1
fl, sl, ʃl	2
pl, tr, kr	3

Table 5.3.: Sonority distances between C1 and C2

Analysis of misperceptions and confusion matrices

To test hypothesis 2, frequencies of reported clusters were compared to those of the target clusters. First, the data were subsetted to include only error trials in which the reported cluster constitutes a legal German CC or CCC cluster in the onset (which also includes clusters not in the set of target clusters); the mean log frequency of the reported clusters was then compared to that of the target clusters. The subset was restricted to consonant clusters as onsets in order to obtain meaningful frequency comparisons. Comparing the target cluster frequencies to the frequency of potentially resulting singleton consonants would be misleading, as single phonemes naturally have higher frequencies (on average) than biphones. For a more direct comparison of target and percept frequencies, it was further determined for each observation in the subset whether the perceptual repair resulted in a higher-frequency or lower-frequency onset cluster.

In parallel fashion, sonority distances—the finer measure as compared to SSP violation—of target and reported clusters were compared to test hypothesis 4 above. Here, the subset of percepts was limited to legal CC clusters (excluding CCC clusters) so that there was exactly one consonant–consonant transition whose sonority distance could be compared to the distance between target C1–C2. The median and mean sonority distances of the reported clusters were compared to those of the target clusters.[13] It was determined for each observation separately whether the perceptual repair led to an improvement or deterioration of the sonority profile.

Additionally, confusion matrices were set up to examine more closely which clusters were commonly confused. Confusion matrices are a very helpful means to discover asymmetries in phoneme (or phoneme sequence) confusions. Here, only test clusters, single C1 or single C2 perception, and perception of the voiced counterpart of the target (as

[13] The median was chosen as the central tendency value because it makes more sense from a theoretical perspective given the fixed 0.5 to 1 step scale for sonority distances. Comparison of *mean* sonority distances, on the other hand, reveals finer differences, although the numbers show distances not actually possible given the steps of the sonority scale.

well as a category summing up all other confusions) were included as possible percepts. This was due to the fact that including all answers given in the open set task would have rendered the matrix practically unreadable. On the other hand, confining the matrix exclusively to the test clusters would have failed to reveal some interesting trends in this open-set task.

In addition to a confusion matrix for the whole onset unit, separate confusion matrices for the individual consonants (C1 and C2) were created. The default for assigning positions to the reported consonants was to label the first consonant of a percept as C1, the next as C2 and so on. There were some exceptions, however:

1. If the first consonant of the percept corresponded to the target C2, it was labelled as (reported) C2 and the slot for (reported) C1 stayed empty. For example, if target /tr/ was reported as /r/, this percept was assigned to C2, while C1 was considered to be perceptually deleted.

2. If the second consonant of the percept corresponded to the target C1, it was labelled as (reported) C1 and the first consonant of the percept was assigned to a C0 position. For example, if target /tr/ was reported as /ʃtr/, /t/ and /r/ were labelled as C1 and C2, respectively, and /ʃ/ was considered an addition as C0.

3. If both consonants of the target occurred in the reported percept but there was an additional consonant between them, the first and third consonant of a reported percept were assigned their positions in the target. For example, if target /ʃl/ was reported as /ʃpl/, /ʃ/ was assigned to C1, /l/ to C2 and /p/ was considered an addition in between. This approach was also taken for reported percepts starting with /pf/. For example, if /pl/ is reported as /pfl/, the /f/ is considered an addition after a correctly perceived C1 rather than as part of a C1 substitution.

5.5. Results

The overall recognition of onset clusters was above chance level, with a mean error rate of 27.5%. Performance varied widely, however, both between subjects (error rates ranging from 17.9% to 47.5%, sd = 0.53) and between consonant clusters, with an error rate of 6.8% for the most perceptible cluster (/ʃt/) and one of 66.0% for the hardest cluster to recognise (/ps/). Figure 5.1 shows the mean error rates for the individual clusters across subjects.

Inspection of the data revealed that five stimulus items (/skuːk/, /skoːt/, /kruːf/, /ʃnuːk/, /ʃløːs/) had extremely high intercepts (above 2.0 on the logit scale). As there were no obvious reasons for this deviation and auditory inspection of these stimuli did not show any abnormalities, the data from these stimuli were included in the analysis.

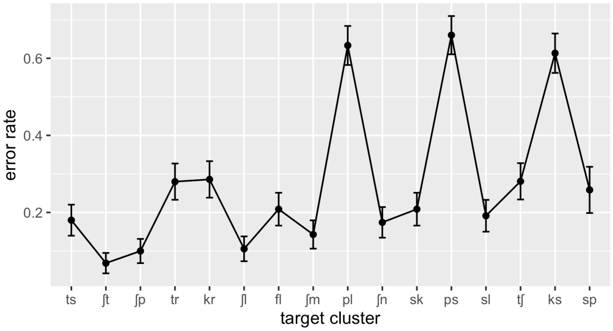


Figure 5.1.: Error rates over consonant clusters in a descending order of type frequency

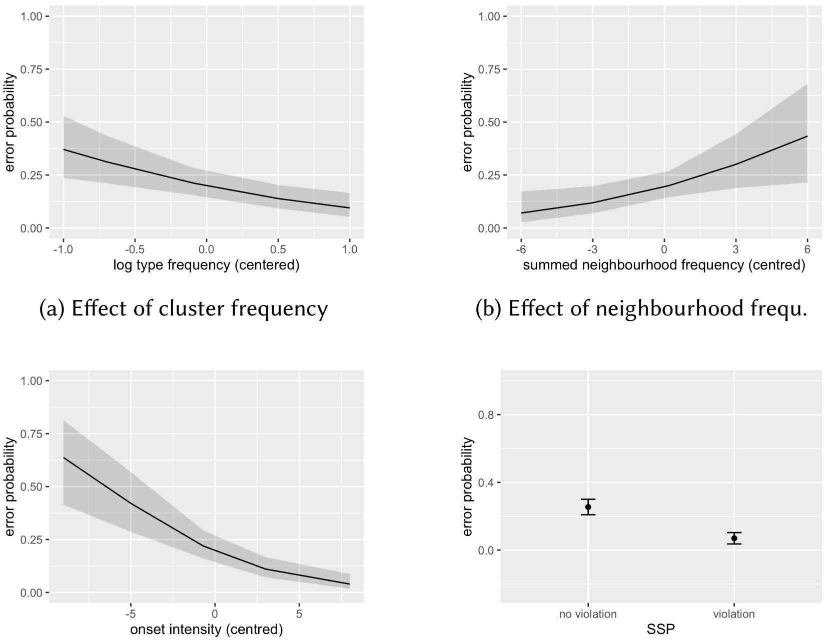
5.5.1. Logistic regression

Log type frequency, neighbourhood frequency, onset intensity, and sonority violation of a consonant cluster had significant effects on its recognition rate, while generalised cluster frequency and salience-based well-formedness did not. Table 5.4 shows the estimates, standard errors and z values for the individual predictors and Figure 5.2 shows the effect plots.

predictor	estimate	SE	z value	
intercept	-0.74804	0.35489	-2.108	*
log cluster frequency	-0.86488	0.26272	-3.292	***
generalised log cluster frequency	-0.00891	0.03090	-0.288	
SSP violation	-1.51393	0.62098	-2.438	*
summed neighbourhood frequency	0.19271	0.08013	2.405	*
onset intensity	-0.22079	0.04696	-4.701	***
salience-based well-formedness	-0.43696	0.50741	-0.861	

Table 5.4.: Model output of the best-fitting model
Significance codes: *** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1
Formula: `error~logFreq + son.vio + accNF + salience + logFreqGen + ons.intensity + (logFreq + son.vio + accNF + salience + ons.intensity|subjID) + (1|onset.targ/stimulus)`

As can be seen, the effect of log cluster frequency is as hypothesised: the higher the frequency of a consonant cluster, the lower the error rate in recognising it. Concerning neighbourhood frequency, the effect takes the opposite direction: the higher the frequencies of the neighbour clusters, the higher the error rate of the target cluster. For the control variable of onset intensity, the data also meet the expectations: clusters of higher intensity showed lower error rates. This suggests a perceptual advantage for high-intensity clusters over and above the intensity variation between cluster classes, which is covered by the random effect of cluster, nested under item. SSP violation does not show the hypothesised influence. There is a significant effect of sonority on recognition error rates, but it goes in the opposite direction: consonant clusters that violate the SSP are perceived better than those that do not. A separate model with sonority distance as a predictor instead of SSP violation (all other parameters being the same as in the best model



(a) Effect of cluster frequency

(b) Effect of neighbourhood frequ.

(c) Effect of onset intensity

(d) Effect of SSP violation

Figure 5.2.: Significant effects in the L1 perception experiment (numeric values are centred)

described above) did not converge (max grad = .005). Moreover, in the model output, sonority distance did not show a significant effect.

5.5.2. Analysis of misperceptions and confusion matrices

Analysis of misperceptions

According to hypothesis 2 (Section 5.3), perceptual repairs should mainly result in HF consonant clusters. A look at the frequency distribution of the reported clusters[14] supports this hypothesis (Figure 5.3). By far the most common outcome of a misperception is the onset with the highest frequency (3.24 on a log scale), /ts/. In general, a trend can be identified for the number of observations (i.e., false positives) to increase with the frequency of a cluster.

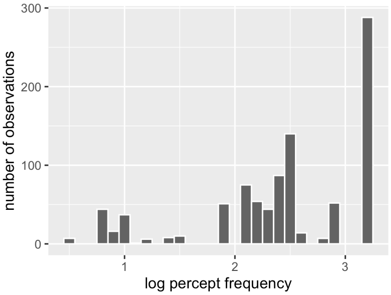
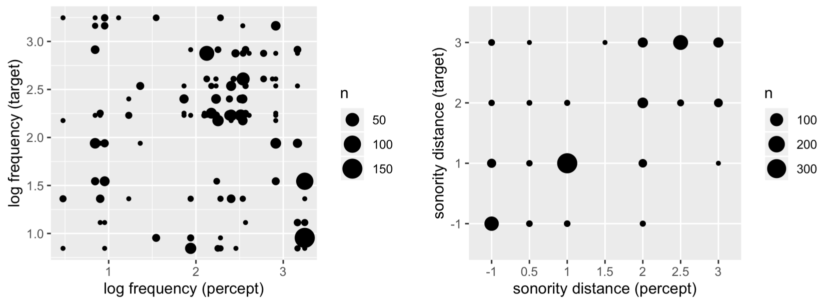


Figure 5.3.: Log frequencies of reported clusters in misperceptions

However, more relevant than the absolute frequencies of the reported clusters is a comparison between target and reported cluster frequencies. The hypothesis was that misperceptions more often than not result in a higher-frequency cluster than the target. Figure 5.4a displays the frequency of the target cluster compared to that of the reported cluster for all error trials in which the percept constitutes a legal German CC or CCC cluster.

As can be seen, there are far more observations of HF percepts (more and bigger dots in the right half of the plot), while the frequencies of

[14] This is based on the data subset including only misperceptions with legal CC or CC outcomes.



(a) Log frequencies

(b) Sonority distances

Figure 5.4.: Comparison of target and percept characteristics in misperceptions (Dot size represents the number of observations.)
Note that in 5.4b the scale is finer for percepts because both stops and fricatives were subdivided into voiceless and voiced variants and the voiced variant was given a value 0.5 points above the voiceless one. Percepts included clusters such as /dr/ = distance of 2.5, while the set of target clusters did not contain any voiced obstruents.

the targets are relatively evenly distributed across the height of the plot as predefined by the design of the experiment. So HF consonant clusters do attract responses from lower-frequency targets. Numbers confirm this visual impression: the mean log frequency of reported clusters lies above that of presented clusters (2.45 vs. 1.92, respectively). Of the 1,498 misperceptions in the response data, 559 perceptual corrections resulted in a consonant cluster of a higher frequency than the target, whereas only 334 corrections resulted in a lower-frequency cluster. (The remaining cases are responses where the onset was not a legal CC or CCC cluster and therefore, the frequencies of target and response were not compared.) So the number of misperceptions directed towards higher frequency is almost twice as high as that directed at lower frequency. This means that the direction of misperceptions constitutes a strong trend but not an absolute rule. This is in accordance with the hypothesis. Importantly, however, whether a target cluster was reported as a higher-frequency cluster or not depended on its own frequency. The HF clusters (/ts/, /ʃt/, /ʃp/, /tr/, /kr/) were perceptually "repaired" to higher-frequency clusters in only around 7% of all cases in which they were misperceived, while the clusters in the medium frequency range (/ʃl/, /fl/, /ʃm/, /pl/, /ʃn/) were repaired in about 64% and the LF clusters (/sk/, /ps/, /sl/, /tʃ/, /ks/, /sp/) in a full 87% of cases.

As far as sonority distances are concerned, it was expected that misperceptions should improve the sonority profile of the cluster and therefore tend to result in a cluster of a higher sonority distance (hypothesis 4). As can be seen in Figure 5.4b, most misperceptions preserved the sonority distance between the two consonants in a cluster: observations accumulate where the values for target and percept match. Furthermore, the dots are distributed relatively evenly over the whole array. The majority of cases where sonority distance of target and percept are not identical are observations where the target cluster has a sonority distance of 3 and the percept a distance of 2.5. They show misperceptions of the stop-liquid clusters /tr/, /kr/, and /pl/ as their voiced counterparts /dr/, /gr/, and /bl/. Hence there is no trend in misperceptions to increase the sonority distance in a cluster.

Actually, of the 1,498 misperceptions in the response data, only 64 improved the sonority profile of the onset, while 238 deteriorated it. Moreover, the median sonority distance of target clusters is not different from that of reported clusters (both 1). The mean distance of the targets is 1.43 and that of percepts 1.15. Here, it can be seen that on average misperceptions slightly decreased (i.e., deteriorated) the sonority distance between C1 and C2.

Confusion matrix

For a more complete picture of the characteristics of the percepts, it is helpful to have a look at the confusion matrix for the response data in the experiment (Table 5.5). Like Figure 5.1, the confusion matrix shows the wide variety in error rates for the individual consonant clusters. For example, it shows that three clusters, /ps/, /ks/, and /pl/ were misperceived more often than identified correctly. In contrast to the graph in Figure 5.1, the confusion matrix also captures which confusions contribute most to the error rates of the individual clusters. So it reveals that in all three cases, misperceptions cumulate on a specific competitor, namely /ts/ for /ks/ and /ps/, and reduction to /l/ (with deletion of C1) for /pl/. So the confusion matrix also reveals the number of false positives (responses for that cluster when the stimulus was a different cluster) for each target onset. For example, /ʃl/ was given as a response approximately as often as it appeared as a stimulus (around 90% being correct identifications and 10% false positives), while /ts/ was given as a response about 1.5 times as often as it appeared as a stimulus.

The confusion matrix for clusters clearly shows the asymmetries in perceptual confusions between stop + /s/ sequences: while /ts/ is identified correctly most of the time and only reported as /ks/ or /ps/ in a minority of cases (1.4% and 0.9%, respectively), the situation is very different in the reverse case. /ks/ is actually reported as /ts/ more often than correctly, and /ps/ is reported as /ts/ almost as much as identified correctly. /ts/ thus attracts far more responses from its competitor clusters than the other way around.

Table 5.5.: **Confusion matrix for consonant clusters**
Rows show target clusters, columns listeners' responses (order of consonant clusters follows token frequencies); columns *C1* and *C2* show cases in which only one of the component consonants was reported as a simple onset, column *voice* reports voicing errors (e.g., perception of /dr/ for target /tr/) and column *other* sums up all the remaining confusions; the sum row shows the total of responses (correct and false positives) each stimulus cluster received, with numbers indicating the percentage of responses in relation to the number presentations; note that the value for /sp/ > /ʃp/ confusions is missing because these cases had to be excluded from the analysis

	ts	**ʃt**	**ʃp**	**tr**	**kr**	**ʃl**	**fl**	**ʃm**	**pl**	**ʃn**	**sk**	**ps**	**sl**	**tʃ**	**ks**	**sp**	**C1**	**C2**	**voice**	**other**
ts	82	0	0	0	0	0	0	0	0	0	0	0.9	0	0.3	1.4	0.3	0	12	0	3.1
ʃt	0	92.6	4.3	0	0	0.3	0	0	0	0	0	0	0	0	0	0.6	0	0	0.6	1.7
ʃp	0	1.7	90	0	0	0	0	0.9	0	0	0.3	0	0	0	0	2.9	0	0	0	4.3
tr	0	0.3	0.3	72	0.3	0	0	0	0	0	0	0	0	0	0	0	1.1	0.9	19.1	6
kr	0	0	0.3	0.3	71.4	0	0	0	0	0	0	0	0	0	0	0	3.1	5.4	12.9	6.6
ʃl	0	0.3	0.3	0	0	89.4	5.1	0	0.3	0.6	0	0	2	0	0	0	0	0	0	2
fl	0	0	0	0	0	3.7	78.3	0	3.7	0	0	0	0	0	0	0	0.9	2.6	0.9	10
ʃm	0	0	0.3	0	0	0.6	0.3	85.7	0	6.9	0	0	0	0	0	0	0.6	1.7	0	4
pl	0	0	0.3	0	0.3	2	8.6	0	36.7	0	0.3	0	0	0	0	0.3	0.9	18.3	6	26.4
ʃn	0	0	0.3	0	0	4.3	0.3	8	0	82.6	0	0	0	0	0	0	0.9	2	0	1.7
sk	0	3.7	6.6	0	0	0	0	0	0	0	78.6	0	0.3	0	0.3	6.3	0.3	0	0.6	3.4
ps	28.9	0	2	0	0	0	0	0	0	0	0	33.7	0	0	4.9	2.3	0	22.3	0.3	5.7
sl	0.3	0	0	0	0	0.6	2.9	0	0	0	0.6	0	80.9	0	0	0	0	0	0	14.9
tʃ	1.1	1.4	0	0	0	0	0	0	0	0	0	0	0	71.9	0.3	0	9.5	15.2	0	0.6
ks	43.8	0	0	0	0	0	0	0	0	0	1.1	2	0	0	36.4	0	0	10.9	2.3	3.4
sp	0.5	1	–	0	0	0	0	0	0	0	13.2	0	0	0	0	72.7	0	1	1.5	10.2
sum	156.6	101.0	104.7	72.3	72.0	100.9	95.5	94.6	40.7	90.1	94.1	36.6	83.2	72.2	43.3	85.4				

Table 5.6.: Confusion matrix for C1

	–	p	t	k	b	d	g	h	f	s	ʃ	v
p	20.7	39.2	14.7	8.2	4.0	0.9	1.7	1.0	5.4	0.7	1.3	0.9
t	9.3	0.7	81.3	0.7	0	6.4	0.3	0.1	0.2	0.6	0.4	0
k	8.4	1.6	22.7	56.7	0.3	0.4	8.0	0.6	0.7	0.1	0.3	0
f	2.9	4.0	0	2.9	1.1	0	0.9	0	83.1	0	3.7	0.9
s	0.4	0.1	0.1	0.1	0	0	0	0	1.3	92.8	5.0	0
ʃ	0.7	0.1	0	0.1	0.2	0	0	0	1.3	1.8	95.4	0

Table 5.7.: Confusion matrix for C2

	–	p	t	k	b	d	g	f	s	ʃ	v	m	n	l	r
p	2.7	87.2	1.8	5.0	0.7	0	0.2	0	0	0	1.3	0.5	0.2	0	0.4
t	0	4.9	94.0	0.3	0	0.6	0	0	0	0	0	0	0	0.3	0
k	0.9	12.9	5.7	79.4	0	0	0.6	0	0	0	0	0.3	0	0.3	0
s	3.4	0.3	0	0	0	0	0	0.3	95.6	0.1	0.2	0	0.1	0	0
ʃ	10.9	0	0	0	0	0	0	0	1.4	87.1	0.3	0.3	0	0	0
m	0.9	0.6	0	0	0	0	0	0.3	0	0	0.9	89.1	7.1	1.1	0
n	1.1	0.3	0	0	0	0	0	0	0	0	0	8.0	86.0	4.6	0
l	4.1	0.1	0.1	0.2	0	0	0	0	0	0	0.3	0.8	0.6	93.6	0.2
r	3.7	0.3	0	0	0	0	0	0	0	0	0.1	0	0	0.1	95.7

It can also be seen that while for most onset clusters a competing cluster is the most common confusion, three clusters (/ts/, /pl/ and /tʃ/) are more often reduced to C2 than reported as another cluster. Strikingly, there are generally far fewer reductions of a cluster to C1 (97 cases) than to C2 (321 cases) in the data, and this can mainly be attributed to stop-initial clusters. Another difference between the clusters revealed by the confusion matrix is that /tr/ and /kr/ (and partly /pl/) produce many voicing errors, while the other clusters do not.

5.6. Discussion

It was assumed that both usage-based factors, such as cluster frequency and neighbourhoods, and the theoretical language-structuring principle of sonority influence the perception of syllable-initial consonant clusters. The hypotheses were that HF clusters would be recognised better than LF clusters, and clusters that conform to the SSP better than clusters that violate it, and that frequency has a greater influence on recognition than sonority sequencing. It was also expected that acoustic factors would influence perception. The results of the experiment show that both acoustic and usage-based factors influence perception of pseudowords in noise. The theoretical concept of sonority, on the other hand, is not supported by the recognition data. The results for the individual predictors and their interpretations will now be discussed in turn.

5.6.1. Acoustic factors

The present study demonstrated the influence of acoustic factors on speech perception. There was a significant effect of onset intensity on error rates. As was to be expected, the higher the intensity of an onset cluster, the more reliably it was recognised.

In contrast to onset intensity, onset duration did not have a significant effect on recognition rates. However, in a model set up post hoc and including only acoustic parameters (see Table A.2 in the Appendix), it did yield a significant effect, which was even bigger than that of onset intensity in effect size. This means that an actual effect of onset duration might have been masked by other, correlated, parameters. Testing for correlations between onset duration and the parameters of the model reported in Table 5.4 (p. 109) reveals that three of them are relatively highly correlated with onset duration (salience-based well-formedness: r_{pb} = .65, p < .001; SSP violation: r_{pb} = .59, p < .001; generalised cluster frequency: Pearson's r = .54, p < .001; for all other predictors Pearson's r < .30). All of them divide the onset clusters into classes (that is true even for the generalised frequencies because these values basically assign clusters to their natural classes). So the lack of a duration effect is not surprising, considering that the variation in duration between onset cluster categories (such as natural classes or SSP-violating vs. SSP-conforming clusters) is naturally much larger than that within one category and the former is already accounted for by the structural predictors in the optimal model. It can therefore not be concluded that onset duration does not have an influence on recognition rates simply from the absence of a significant effect in modelling the response data.

The results of the experiment also largely replicated earlier studies showing the differential noise-resistance of various classes of phonemes. Specifically, they sustained that sibilants are very noise-resistant while stops are much more likely to be lost. As can be seen in Figure 5.5, stop-initial clusters have the highest error probabilities; they are approximately twice as high as those of the other consonant clusters. A glance at the confusion matrices confirms that their high error rates are due to the initial stop being misperceived (or perceptually deleted). Sibilant-initial clusters, on the other hand, have the lowest error probabilities and the confusion matrices (Tables 5.6 and 5.7) reveal that the sibilants themselves have very high recognition rates mostly above 90 %.

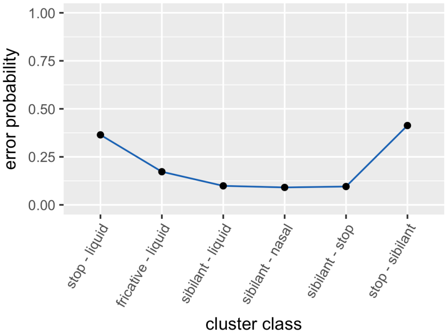


Figure 5.5.: Error rates over classes of consonant clusters

This difference between the two classes is easily explained in terms of their acoustic properties. Sibilants are characterised by high-frequency energy and high intensity (e.g., Reetz & Jongman, 2009; Wright, 2001). And as Moreno-Torres et al. (2017; p. 3080) summarise, “the presence of energy above the masking noise is [probably] the most important predictor of resistance”. It is likely that the multi-talker babble did not mask the sibilants “sufficiently”. A comparison of the specific audio files used in the experiment (i.e., a comparison of the sibilant parts of the stimuli with the multi-talker babble) supports this view (see Figure 5.6, a random sample of eight babble files and eight sibilants from different phonetic contexts in the set of stimuli). All of the randomly selected sibilants exceed the babble in amplitude in the frequency spectrum of 7–8 kHz, many of them even for a larger frequency area. Moreover, fricatives are also long in duration, which gives listeners more time to identify their acoustic cues correctly. In general, the longer in duration and the more intense a signal is, the better internal cues it has (cf. Wright, 2001). That puts fricatives and especially sibilants at an advantage.

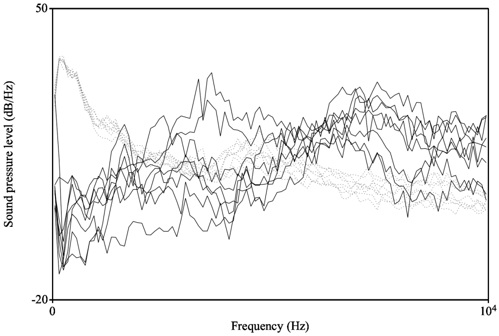


Figure 5.6.: Comparison of LTAS of randomly selected babble (grey dotted lines) and sibilant tokens (/s/ and /ʃ/ from different onset clusters, black solid lines)

Stops, on the other hand, are very short in duration (so their acoustic cues are more transient and thus harder to catch) and have a lower amplitude.

In the present experiment, the inequality in noise-resistance between the different phoneme classes is even greater than in many other perception studies because of the phonetic context. Many of the studies reported—and all of the aforementioned ones that the similarity index is based on—tested perception of onset consonants in a CV context. This means the formant transition to the vowel provided some additional information as to consonant identity. This cue is especially important in the case of stop consonants, which have poor internal cues. Their release bursts are transient cues which are not only more easily missed but also more likely to be masked by abrupt changes in environmental sounds (Wright, 2001; p. 256). In the present experiment, stops in C1 position lacked the formant transition to a vowel and thus had at best external cues in the formant transition carried by a liquid or no external cues (if followed by a fricative). Consequently, stop recognition was aggravated by the two main cues—formant transition and burst—being obscured. How detrimental this is, is revealed by the confusion matrices: the most common perception error throughout the experiment was misperceiving a stop in C1 position (/tr/, /kr/, /ps/, /ks/) or not perceiving it at all (/ts/, /tʃ, /pl/). Hence, the internal cues to stop consonants were obviously not sufficient for listeners to perceive them reliably in many cases, and the following consonants—sibilants or liquids—proved to be poor bearers of external cues, the latter in spite of their formant structure. In fact, the data suggest that /l/ is hardly a better carrier of the cues to a preceding stop than /s/ (cf. the recognition rates of /pl/ [36%] vs. /ps/ [33%]).

As reported in the Results section, C1 is perceptually deleted far more often than C2, and this trend is mainly due to stops in C1 position being deleted. A comparison of error rates for the same stop consonant in C1 and in C2 position supports this interpretation. Stops

consonant	C1	C2
p	39.2	87.2
t	81.3	94.0
k	56.7	79.4
s	92.8	95.6
ʃ	95.4	87.1

Table 5.8.: Recognition rates (in %) of stops and sibilants as a function of their position

are recognised considerably better in C2 position than in C1 position. The same is not true for sibilants, the other class of phonemes that appears in both positions in the set of consonant clusters used here (see Table 5.8). They are recognised approximately equally well in both positions, /ʃ/ even a bit better in C1 position. The reason for this discrepancy between stops and sibilants is most likely two-fold. Firstly, the highly transient cues to stop identity are very difficult to perceive, especially with the sudden onset of the stimulus after 450 ms of pure babble. Secondly, the absence of supporting formant transitions to the vowel, which the stops heavily depend on, impedes perception further. However, the high error rate of /pl/ (with /p/ being either perceptually deleted or perceived as /f/ or /ʃ/) indicates that a following phoneme with a formant structure is not enough to secure recognition.

Cue robustness can also explain why the most common error for /ts/, /pl/, and /tʃ/ is reduction to C2, whereas for the other onset clusters a competing cluster is the most common confusion. All three clusters start with a stop, which is easily missed because of its short duration and reduced burst when followed by another consonant, and have either a strident with their high energy or /l/ with its clear formant structure as C2, all sounds that are relatively noise-resistant.

Returning to the cue-based evaluation of consonant sequences in Table 4.1 (p. 58), it becomes clear that this classification makes relatively good predictions concerning error rates (cf. Figure 5.1 on p. 108). On the whole, the optimal sibilant-initial clusters have very low error rates, followed by clusters of medium cue robustness (obstruent–liquid). Finally, the stop–sibilant clusters with poor cue robustness on average have the highest error rates. There are three notable exceptions, however. The error rates of /ts/ and /tʃ/ are lower than would be expected for clusters with poor cue robustness, and those of /pl/ are higher than expected for medium cues robustness. The good recognisability of /ts/ could be attributed to its high frequency (see Section 5.6.2 below). Nevertheless, the fact that the error rates for the LF cluster /tʃ/ are also remarkably low leaves room for speculation that it might have to do with the specific make-up of these two clusters. Homorganic clusters differ from heterorganic ones in that they share the cues to place of articulation. There is, however, hardly any research concerning the consequences for perceptibility. Even regarding the cross-linguistic preferability of homorganic over heterorganic clusters, which could provide some clues to perceptibility and ensuing persistence in language change, there is no consensus among researchers. While some claim that homorganic clusters tend to be avoided (a variant of the Obligatory Contour Principle, e.g., Greenberg, 1965), others maintain that homorganic clusters are *preferred* over heterorganic ones (Hume, 2003; Jun, 1995). Homorganic consonant–liquid sequences share the vowel formant transitions (Ali & Van Heuven, 2009), which could create stability in perception. For stop–fricative clusters, however, no such advantage is immediately apparent. Here, it seems more likely that the shared place of articulation makes the stop's cues more vulnerable to being drowned out in the following friction noise. Indeed, Baroni (2014) found that initial plateau clusters differing in place of articulation are identified more easily than homorganic ones. However, /tʃ/ reached a reasonably high recognition rate in his experiment as well (95.3%, with the experiment mean for initial clusters being 83%). Moreover, as will be argued below, a first tentative comparison with /pf/ gives no indication that homorganicity of stop–fricative structures generally facilitates recognition. It thus remains puzzling what exactly gives this cluster a perceptual advantage.

The stop /p/, on the other hand, seems to be an exception within its class, with consistently high error rates. It has proven difficult to perceive both in /pl/ and /ps/ onsets and has a lower recognition rate than the other two stops in C1 position, despite /pl/ onsets being both longer in duration and higher in intensity than /tr/ and /kr/, see Figures A.1b and A.1a in the Appendix). It is unlikely that the difference in recognition rates between these three clusters is due to the consonant in C2 position. [l] with its clear formant structure should be a better carrier of the preceding stop's external cues than [ʁ], which is mostly realised as a fricative. Furthermore, post-hoc analysis of filler items with initial /p/, /pr/, and /pf/ revealed that the perception of /p/ in these onsets was equally impeded. Therefore, the difference between the three clusters seems to lie in the stops themselves. This is in accordance with previous research. For example, Moreno-Torres et al. (2017) note that in their study as well as in previous ones, frontal consonants are least resistant to noise.[15] So there seems to be some acoustically-based variation within a manner class as well. Apart from these three exceptions, though, the cue-based cluster categorisation predicts the error rates in the experiment well.

The data therefore suggest that the robustness of acoustic cues is one of the most decisive factors for correct consonant identification. If the internal cues are robust, as in the case of sibilants, the recognition rate is high. However, if the internal cues to consonant identity are weak, as in the case of stops, the recognition is highly dependent on a favourable neighbouring segment, i.e. one that is a good carrier of its external cues. The stops in the sibilant–stop sequences were recognised better than the ones in stop–sibilant sequences because their

[15] Wright (2001; p. 264), on the other hand, found that "labial place of articulation is the most reliably recovered" but also mentions that his results deviate from Hume et al. (1999) and Jun (1995).

cues were carried well by the flanking vowel. This finding is in line with current research. According to Henke et al. (2012; p. 77), the most important aspects for advantageous sound sequencing are modulation (both in terms of amplitude and frequency) and "how good a carrier of transitional cues a sound is for flanking segments". While the role of modulation was not tested explicitly here, the importance of neighbouring segments for transitional cues was corroborated.

However, the exceptions noted above show that cue-based explanations cannot account for the whole range of variability in the data. In the following section, it will be discussed to what extent cluster frequencies provide explanations for the cases not captured by acoustic ones and which general patterns in the data can be attributed to an influence of cluster frequencies on recognition.

5.6.2. Frequency

It was hypothesised that the frequency of a consonant cluster would affect its recognition in noise. The regression analysis indeed revealed a significant effect of cluster frequency. The higher the frequency of a cluster, the fewer errors occurred in its perception. Here, some prominent cases will be discussed for illustrative purposes, the origin of the effect debated, and conclusions with regard to the relevance of sublexical frequencies in speech perception drawn.

The frequency effect becomes especially evident where the acoustic characteristics of the consonants fail to explain recognition rates, namely when clusters of similar composition have divergent recognition rates or when the same clusters show divergent recognition rates depending on the listener group. Both can be observed here. For example, error rates for HF /ʃt/ and /ʃp/ (6.9% and 10%) are the lowest in this study. In contrast, the phonetically very similar /sp/ at the other end of the frequency scale has an error rate of 25.9%—a multiple of the HF clusters' error rates. The fourth sibilant–stop cluster also fits into the picture: it ranks between the extremes both in terms of frequency and error rates.

It is important to exclude the possibility that this difference in error rates is caused by acoustic differences between the sibilants. As the vast majority of perception errors occurred on the stop, not the sibilant[16], and there is no reason to assume that the stop's perception was significantly influenced by a *preceding* sibilant, an explanation based on the acoustic characteristics of the sibilants is unlikely. However, more importantly, there is a strong discrepancy between the present results and the error rates for the same clusters in a comparable[17] perception study with Italian and Dutch listeners (Baroni, 2014). In the latter, /ʃt/ and /ʃp/, which are both phonotactically illegal for the listeners, show relatively high error rates (56% and 38% respectively). Conversely, /sp/ and /sk/, which are legal in Italian and Dutch, had very low error rates (1.6% and 4.7%, respectively) in Baroni's study. These divergent results in the two studies can best be explained by the listeners' phonotactic knowledge influencing their perception. While the phonetic characteristics of the clusters are the same in both experiments, their phonotactic status is different and this difference is mirrored in the error rates. That means that under the adverse listening conditions that the listeners faced, they were biased by their structural knowledge to perceive the consonant clusters that are most expectable.

Baroni (2014) interprets the results of his study in terms of absolute phonotactics, differentiating merely between legal and illegal clusters. However, the error rates of the present experiment show the gradience of the effect: the HF clusters /ʃt/ and /ʃp/ have the lowest error rates among the sibilant–stop clusters, LF /sp/ has the highest error rate, and /sk/ ranks in between the two, just as its intermediate frequency rank

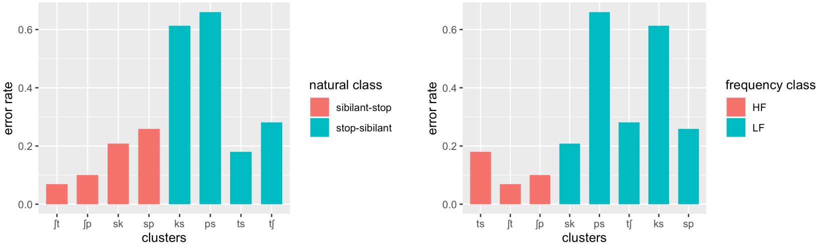
[16] Except for the /sp/ > /ʃp/ cases that had to be excluded from the analysis due to potential spelling confusion.

[17] Note, however, that Baroni (2014) used white noise to mask his stimuli, while multitalker babble multi-talker babble was used in the present study.

would lead one to expect. Statistically, the gradience is underpinned by the significant effect of the numeric predictor frequency. The German preference for /ʃ/-initial clusters is also reflected in the low error rate of /ʃl/ as compared to /sl/. Again, that corresponds to the lexical distribution of these clusters. In Dutch or English listeners, the opposite pattern would be expected.

Similarly, the strong divergence in stop–sibilant error rates (61% and 66% for /ks/ and /ps/, respectively, and 18% for /ts/) corresponds remarkably well to their frequency difference. It could be argued that the extraordinarily good perceptibility of /t͡s/ is due to its segmental status as an affricate, which distinguishes it from /ks/ and /ps/, rather than its high frequency of use as a German onset. Acoustically, there are some differences between /t͡s/ and /ts/ such as the rise time (e.g. Howell & Rosen, 1983; Mitani et al., 2006). Unfortunately, there is no possibility of controlling for this confound within the scope of the present study. A possible approach would be to compare the perceptibility of /t͡s/ to that of /p͡f /, whose CELEX type frequency is about a tenth that of /t͡s/. A first tentative impression can be obtained by looking at results for filler items with /p͡f / onset in the present experiment. The two filler stimuli beginning with /p͡f / (summing up to 70 observations across participants) showed an error probability of 80%. This suggests no advantage for affricates over true consonant clusters. However, participants might have been biased against responding <pf> due to the scarcity of this stimulus onset in the set of stimuli, so that any conclusions drawn should be viewed very cautiously.

Concerning error rates, not only does /ts/ group better with HF clusters than with stop–fricative clusters (which share certain phonological–phonetic features) but /sp/ also groups better with LF than with fricative–stop clusters, even with the doubtful cases of <sp> transcription excluded. All in all, grouping the eight obstruent–obstruent clusters according to frequency leads to a more homogeneous distribution of error rates within groups (in that one of the groups, HF clusters, has consistently lower error rates than the other) than grouping them according to manner of C1 and C2 (see Figure 5.7), although taking both factors into account would map the error rates most reliably.



(a) Grouping according to natural classes (b) Grouping according to frequ. classes

Figure 5.7.: Error rates of obstruent–obstruent clusters

These observations show that not only a cluster's natural class with its connection to acoustic cues but also its frequency in the language influences its recognition in perception.

Turning to the outcomes of misperceptions now, it is obvious that frequency plays a role here, too. Generally speaking, the higher the frequency of a cluster, the more false positives it showed in the experiment (see Figure 5.3 on p. 111). The comparison of target and percept frequencies in misperceptions showed that the reported clusters on average have a higher frequency than the targets and that especially LF clusters are perceptually repaired to clusters of a higher frequency.

That was especially the case for /ts/ (the onset with the highest type frequency in the test set), which attracted a high number of responses from the two LF target clusters /ps/ and /ks/. Target /ts/, on the other hand, is hardly ever perceptually repaired to /ps/ or /ks/, so the confusion is asymmetric. For example, the number of /ks/ > /ts/ confusions is more than 30 times higher than that of /ts/ > /ks/ confusions. It is also worth noting that HF /kr/ was not perceptually repaired to higher-frequency /tr/ very often. Hence, in phonologically similar pairs of consonant clusters, the direction of confusion is clearly biased towards the HF cluster in the present study and this bias is strongest for HF clusters. That even leads to perceptual asymmetries between two phonemes being reversed when the frequency relations of the clusters that they appear in are reversed. An example will illustrate this point. As discussed above, the perceptual illusion /s/ > /ʃ/ is very common in sC clusters, while the opposite is not true. For /ts/–/tʃ/, on the other hand, the confusion goes in the opposite direction (/ʃ/ > /s/), again turning a LF cluster into a HF one. Asymmetries in perceptual confusions have been observed before and have been attributed to acoustic–phonetic factors, such as the energy profile of the consonants involved (Chang et al., 2001; Moreno-Torres et al., 2017) and the phonetic context (Woods et al., 2010), to phonological factors, such as phonological underspecification of phonemes (Lahiri & Reetz, 2002, 2010), and to higher-level factors such as phoneme frequency (Benkí, 2003; Moreno-Torres et al., 2017), lexical frequency (Benkí, 2002), and phonological neighbourhoods (Benkí, 2002). All in all, the higher the frequency of a consonant cluster, the more often it was the outcome of a misperception, although there are some outliers. These can mostly be explained by the availability of neighbours, as will be discussed below.

These findings support the hypothesis that speech perception is guided by the listeners' phonotactic knowledge. Both parts of the hypothesis are borne out by the data: 1) the more frequent a consonant cluster is, the more likely it is to be perceived correctly, and 2) less frequent clusters tend to be misperceived as (similar) clusters that are higher in frequency.

In conclusion, the tendency is for listeners to perceptually repair LF clusters as higher-frequency clusters. This is not surprising, considering the generally established frequency effect in psycholinguistics and earlier findings discussed in Section 5.2.4. However, onset clusters have hardly been considered as a unit to which this principle applies.

As described earlier, most studies concerned with consonant cluster phonotactics regard it as a categorical measure and find a perceptual advantage for legal over illegal clusters (and the tendency for perceptual correction of the latter) but no gradient advantage for HF clusters. For example, in describing the challenges for a theory of a phonotactic grammar, Lentz (2011; p. 35) states that "[i]f it predicts perceptual illusions for illegal sound combinations, because the illusion is more probable, it also predicts illusions for marginally legal sound combinations when there are very well-formed alternatives." He mentions this as a problem for theories of probabilistic knowledge of phonotactics, but this is exactly what can be seen in the present data. The participants experienced perceptual illusions—not in the form of epenthetic vowels, as often induced by studies on illegal clusters, but in the form of cluster confusions and reductions—for less well-formed clusters, just as participants in previous studies experienced perceptual illusions for phonotactically ill-formed clusters. We can therefore conclude that phonotactic knowledge is actually gradient, or at least that a gradient form of phonotactic knowledge exists (possibly alongside a categorical one).

That raises the question of what this gradient phonotactic knowledge is based on. As mentioned before, (log) type frequencies of the clusters were used here because they have turned out to make the best predictions concerning speech processing. They showed the hypothesised frequency effect reported here. But in order to examine the role of token frequencies in consonant cluster perception as well, an additional model was set up, identical to the one reported but with token frequencies (taken from the CLEARPOND corpus, Marian et al., 2012) instead of type frequencies.[18] There, token frequencies showed the same effect as type frequencies before, but the variance in error proba-

[18] The model did not converge, but since both the frequency effect and the effects of the other parameters are the same as in the type frequency model, it can be assumed that the output is nonetheless relatively reliable.

bilities at the HF end of the spectrum was bigger. As the two frequency measures are highly correlated (Pearson's r = .812, p < .001), the design of this study does not allow to draw reliable conclusions as to which of them is the cause of the effect. The lower standard error of the type frequencies, however, speaks in their favour. The conclusion drawn by Brysbaert et al. (2011) that "Celex frequencies [...] have had their best time" is not supported by the present data, either. The CELEX type frequencies proved to be an adequate measure that captures the listeners' psychological reality reasonably well, as can be seen by the significant effect in the hypothesised direction. Frequencies based on television subtitles (i.e., the CLEARPOND frequencies), as proposed by above-named authors without distinction between type and token frequencies, certainly do not fare any better. However, for a direct comparison with type frequencies from a different source, a model identical to the best-fitting one reported but with type frequencies taken from the elexiko online dictionary ("elexiko," 2003) instead of CELEX type frequencies was set up post-hoc. The results for all predictors, including type frequencies, remained the same as reported above. A comparison of goodness-of-fit of the models shows a slight advantage for the CELEX model (elexiko: AIC = 4946.8, BIC = 5144.9; CELEX: AIC = 4944.9, BIC = 5143.0). So CELEX type frequencies are not inferior to type frequencies from a more recently (2003 ff.) compiled dictionary.

It also became clear from the study that it is specific, not generalised, frequencies that serve as a resource for phonotactic knowledge (see the lack of a significant effect of generalised cluster frequencies). Language users seem to "track" specific phoneme sequences and not generalise this knowledge to similar clusters. Clusters like /sp/ and /sk/ do not seem to profit from the high frequency of /ʃt/ and /ʃp/ in the German language. Here, the results from the perception experiment deviate from those of learning studies and experiments using acceptability ratings, where phonotactic generalisations over natural classes have been proven to be relevant (e.g., Albright, 2007b; Linzen & Gallagher, 2014). This discrepancy of relevance follows naturally from the different functionalities of frequencies in learning and perception: In learning and acceptability ratings, new material has to be assessed, so generalising from known to unknown sequences of segments is beneficial. In adult L1 perception, on the other hand, the frequency distributions of the presented material in the language are known and can be used directly to guide processing. Generalisation over natural classes would actually distort the phonotactic knowledge and undermine the mechanism that puts frequent phoneme sequences at an advantage. Speech perception is attuned to the frequency distributions of a particular language, where sequences of phonemes of the same classes are not distributed evenly and accidental gaps exist. Generalisations can give a rough idea, for example, that initial stop–stop or stop–nasal sequences are not permitted in the native German phonotactic system and therefore are limited to a few loanwords, but specific cluster frequencies make better predictions to guide perception, preventing for example /ts/ from being perceived as /ks/ too often, although the two are structurally and auditorily similar and could easily be confused. In normal speech perception, higher-level information such as word frequency and semantic context of course comes into play as well, but even there, sublexical frequencies can serve as a first clue, especially in the case of word onsets (Pylkkänen et al., 2002; van der Lugt, 2001). In perception, cluster classes do play a role, however, insofar that their members share acoustic cues, which are highly relevant for perception, leading, for example, to a recognition rate for /sl/ comparable to that of /ʃl/ and /fl/ and far beyond what would have been expected based purely on this cluster's type frequency.

The frequency effect for consonant clusters observed in this study thus meets the expectations based on previous research on the role of phonotactics and sublexical frequencies in speech perception. There are also studies, however, whose results diverge from the present findings. For example, in a gating experiment with all legal Dutch diphones, Warner et al. (2005; p. 70) found only a weak influence of phoneme frequency on listeners' responses and no significant influence of transitional probabilities between the two phonemes and conclude:

> These results suggest that listeners can do quite well at speech perception, and at recognizing individual sounds, from bottom-up information alone. Listeners certainly do not have to rely on higher-level information such as overall frequency or transitional probabilities in order to decide what sounds they are hearing.

While it is certainly true that listeners *can* recognize phoneme sequences from the bottom-up signal alone in quiet listening conditions (as was the case in the gating experiment), the present results show that they can also make use of their statistical knowledge of the language and are very likely to draw on this resource when the bottom-up signal is less reliable.

Finally, it should be noted that one of the clusters tested here, /ks/, is potentially problematic concerning its frequency value. It has been treated as a cluster of very low frequency (with a CELEX type frequency value of 0.95, see also the Figure 3.2 on p. 48 in Chapter 3) and in fact only occurs in very few lexemes like *Xylophon* or *xenophob*, all of them loanwords. However, in southern German dialects, word-initial /ks/ is regularly created through syncope of perfect participle forms (e.g. *gesagt* > [ksa:kt]), leading to numerous word forms with initial [ks] in the spoken language. This process could distort the frequency of this cluster and hence the listeners' experience with it. For this reason, speakers of southern German dialects have been explicitly excluded in the call for participation. However, since it was not possible to limit the study to subjects without any kind of experience with southern German dialects, the extent of their experience with them was gauged by a questionnaire and its effect on the perception of /ks/ analysed separately. (The model summary and effect plot can be found in the

Appendix, Table A.3 and Figure A.3.) The analysis revealed that experience with southern German dialects did not improve perception of /ks/. As a matter of fact, the subjects with the highest familiarity levels with southern German dialects (4 and 5 on a scale from 0–5) had a numerically, but not significantly, higher error probability in /ks/ perception than subjects without that dialect experience. For the other clusters, their error probability was the same as that of subjects without experience with southern German dialects. Moreover, it also turned out that four participants failed to meet the inclusion criterion of not speaking a southern German dialect. They were tested in spite of this failure, but the main regression analysis was repeated without their data to control the results for potential influences of this familiarity. The output of the regression model was the same as with the full data set (with slight deviations in z values) and can be found in Table A.4 in the Appendix. It was therefore considered legitimate to include these participants' data in the analyses and treat /ks/ as a cluster of very low frequency.

This lack of effect of southern German dialect competence is surprising. It can be taken as an indication that [ks] as an underlying cluster and [ks] from underlying /gəs-/ or /gəz-/ have separate entries in the mental lexicon and therefore the assumed low frequency of /ks/ also applies to the listeners with experience with southern German dialects. An alternative interpretation is that these speakers have separate "frequency counts" for different language varieties so that the high frequency of /ks/ in their dialect did not affect its Standard German representation, which was targeted in the experiment.

A similar case is represented by a participant who grew up bilingually with Greek as the other L1. As both /ps/ and /ks/ are far more frequent initially in Greek than in German, this was considered as particularly problematic and his data were inspected extra carefully. They did not deviate from the other participants' data in any way that could be attributed to this circumstance. Both the data from the subjects with dialect competence and the data from the bilingual subject could thus be an indication that sublexical frequencies do not accumulate over different languages or language varieties a speaker–hearer is proficient in, but no definite conclusions can be drawn from data from such a small population sample. The examination of L2 listeners will shed more light on this matter.

To sum up, the results from the perception experiment showed that sublexical frequencies clearly play a role for the perception of pseudowords. Both the recognition rate of the target consonant clusters and the false alarms for competitor clusters are evidently influenced by their respective frequencies as onsets in the German language.

5.6.3. Competition and cluster neighbourhoods

A significant inhibitory effect of neighbourhood frequency was found in the data: the regression model showed that consonant clusters with a high log neighbourhood frequency (i.e., whose phonological neighbours have a high summed frequency) are more prone to perception errors than clusters with a low log neighbourhood frequency (see also Figure 5.2b on p. 110). This is exactly what one would expect in light of activation–competition models of speech perception because clusters with many and highly frequent neighbour clusters have to compete more for recognition. This means the competitive influence of neighbourhoods, found in the work of Vitevitch, Luce, and colleagues (e.g., Vitevitch & Luce, 1998, 1999) and predicted by both PARSYN and ARTPHONE, was also observed in the present experiment. Here, it was not lexical but consonant cluster neighbourhoods, specifically, phoneme sequences derived by changing one phonological feature, that were inspected in order to accommodate for the grain size of the unit studied.

That the inhibitory effect was caused by *cluster* neighbourhoods is a noteworthy finding. Based on theoretical considerations and analogical transfer of partial findings, Vitevitch and Luce (1999) assume that both inhibitory and facilitative effects can occur at any level of processing. However, their study was not designed to test this, and their results only show facilitative effects at the sublexical level (which, as they note, are analogous to word frequency effects) and inhibitory effects of neighbourhood density at the lexical level. The present study can be regarded as support for the assumption of parallel effects on different processing levels. Here, simultaneous effects of target frequency (facilitative) and neighbourhood frequency (inhibitory) were shown at the sublexical level for pseudowords: not only was facilitation (as evidenced by lower error rates) for HF clusters found in response accuracy, but at the same time, consonant clusters in dense and frequent neighbourhoods showed higher confusion rates. Importantly, the effects of cluster frequency and cluster neighbourhood were independent of each other. Correlation between the two measures was very low (Pearson's $r(14) = 0.16$, $p = .55$) and an interaction between them was ruled out during the model fitting process. This shows that both the facilitative effect of high frequency and the inhibitory effect of competition, which have been observed at the lexical level, also exist at a sublexical level. This lends further support to the assumption that both the facilitative effects of frequency and the inhibitory effects of competition can occur at any level of speech processing or, put more generally, that the same mechanisms operate at different levels of processing and on linguistic units of different sizes.

It might be argued that the use of phonological features to derive neighbours does not capture the perceptual reality of the listeners very well. After all, the most common confusions in the response data depended on acoustic cues rather than phonological features. The activation of possible perception candidates—based on the auditory signal—therefore probably also involves acoustic similarity or at least the weighting of phonological features (cf. also Martin & Peperkamp, 2017; for a discussion on the varying importance of phonological features for speech perception). Thus, the unexpectedly high error rates of /tr/, /kr/, and /pl/ (given their frequencies and accumulated neighbourhood frequencies) might also be accounted for. A closer inspection of these clusters shows that they are the only consonant clusters in the sample for which errors in voicing perception would lead to another legal cluster in German (/dr/, /gr/, and /bl/, respectively). Voicing is therefore hypothesised to make a neighbour more confusable.[19] Yet, for now, the measure used—the accumulated frequency of all clusters differing from the target in one phonological feature—was sufficient to show the inhibitory effect of high neighbourhood frequency on consonant cluster perception.

Since misperceptions mostly result in legal phoneme sequences—and even in particularly frequent ones, as the analysis of percepts shows—, it can generally be said that there are not equally many plausible misperceptions for all clusters. For example, from /sk/, only the two very marginal clusters /sp/ and /st/ can be formed by modifying one phonological feature; in contrast, from /pl/, six mainly common clusters can be formed: /kl/, /fl/, /pfl/, /bl/, /pr/, and /ps/. This difference could explain why the error rate for /sk/ is far below expectations, the one for /pl/, on the other hand, far above.

It has also been noted that in contrast to the other clusters /ts/, /pl/, and /tʃ/ are more often reduced to C2 than reported as another cluster. That has been explained in terms of their cue robustness, with the stops being very vulnerable and the sibilants and /l/ quite noise-resistant. However, for /ts/, there is furthermore no phonetically similar HF cluster that easily lends itself to confusion. The closest phonetical neighbour, /ks/, is very low in frequency and, therefore, probably no coequal competitor. It is postulated here that this is the reason why it is both prone to be misperceived itself and rarely the outcome of a misperception. For /tʃ/, on the other hand, a HF confusion candidate would be /ts/, but the two sibilants are most likely phonetically-acoustically too

[19] The confusion matrix confirms that at least for /tr/ and /kr/ a voicing error is in fact the most common confusion. Why /pl/ does not produce more voicing errors is at present unclear.

strong and distinct from one another to be confused. Instead, the most common cluster confusion for /tʃ/ is /ʃt/ (albeit making up only 1.4% of /tʃ/ perceptions), where the components are phonetically (nearly) identical but reversed.[20] In accordance with hypothesis 2, the latter is a cluster of very high frequency of use.

In contrast, the confusion matrix showed that targets /tr/, /kr/, and partly /pl/ led to many voicing errors, while the other clusters do not. That is obviously due to the fact that those three are the only onsets for which a voicing error leads to another legal (or native, if /vl/ is considered legal) German onset cluster.

The above observations seem to suggest that there are several factors relevant to consonant cluster confusions: how strong the cluster is in itself, both acoustically and in terms of frequency of use, and how strong the competition is. Are there phonetically similar clusters in the language inventory, how similar are they to the target, and how frequent are they? All these factors are strongly reminiscent of the PARSYN model of speech perception and its predictions. Recall that in PARSYN, perception is a function of the intelligibility of the stimulus, its discriminability from its competitors (neighbourhood density) and their relative frequencies. The observations are also in line with ARTPHONE. According to ARTPHONE, the learned pathways are an important feature of the recognition process. When some features of the acoustic signal are masked by noise, as in the present experiment, more items which might be in accordance with the auditory input are activated (and their resonances not terminated by mismatch reset), and learned expectations lead to stronger resonances with the higher-frequency items. Therefore, a HF item that does not correspond to the stimulus might lead to a resonant state if the auditory input is ambiguous enough. It is therefore crucial to recognition how strongly

[20] In contrast to the speech production experiment, however, reversal errors were on the whole very rare in the perception experiment, making up only 18 of the 5,452 observations.

masked the auditory signal is, which items are activated because of the ambiguity and how strong the expectations for them are.

In order to further test the relevance of these factors, for each target cluster the strongest competitor—i.e., the consonant cluster it was confused with most often throughout the response data—was determined and a logistic regression model was set up, in which the likelihood of the recognition of the strongest competitor was predicted from its perceptual similarity to the target, the difference in frequency between them, and the neighbourhood density of the target cluster. Table 5.9 shows the main competitor for each target cluster along with their frequency difference and similarity index.

target cluster	competitor	similarity index	frequency difference
fl	pl	5.80	−0.171
kr	gr	6.90	−0.069
ks	ts	3.53	2.292
pl	kl	7.20	0.283
ps	ts	3.32	1.702
sk	sp	6.24	−1.094
sl	fl	1.40	1.041
ʃl	fl	0.27	−0.133
ʃm	ʃn	5.67	−0.077
ʃn	ʃm	10.49	0.077
sp	sk	7.20	1.094
ʃp	sp	0.69	−2.069
ʃt	Sp	1.89	−0.249
tr	dr	1.89	−0.752
ts	ks	2.18	−2.292
tʃ	ʃt	NA	2.049

Table 5.9.: Targets and clusters they are most commonly confused with; perceptual similarity is averaged from perception studies, frequency difference between the two clusters is calculated based in their log frequencies

Perceptual similarity was calculated by averaging the percentage of confusions between the consonants they differ in from a number of studies which investigated the perception of single consonant onsets (Bellanova, 2016; Benkí, 2003; Jürgens et al., 2007; Lecumberri & Cooke, 2006; Marchegiani & Fafoutis, 2015; B. T. Meyer et al., 2010; Moreno-Torres et al., 2017; Woods et al., 2010). For example, to arrive at a similarity index for /ʃm/ and /ʃn/, the mean percentage of trials that /m/ was mistaken for /n/ in the above studies was calculated. The values are unidirectional, so the numbers for /ʃm/ to /ʃn/ and /ʃn/ to /ʃm/ confusions could (and do) differ. The studies (see above) were chosen because they are phonetically oriented and test the consonants under consideration in a CV context. This was meant to serve as a purely perceptual baseline against which contextual effects of cluster frequency could be compared. It has to be noted, however, that the values are merely approximations because the confusability of a consonant varies with its phonetic context, as was described in Section 4.3.

The binary dependent variable in the regression model was the perception of the main competitor instead of the target or any other competitor in each trial. Frequency difference between the target and its main competitor, similarity index between the target and the competitor, and the neighbourhood density of the target were entered into the model as fixed effects. Random slopes for all three predictors by participant and a random intercept for stimulus (nested in target cluster) were added as random effects. The model output can be seen in Table 5.10.

As can be seen, the frequency difference between target and main competitor has a significant influence on the recognition of the competitor such that the more frequent the competitor is in comparison to the target, the more likely it is to be perceived. However, neither the similarity between the target and the competitor cluster nor the neighbourhood density of the target shows significant effects. This is rather unexpected. It could be assumed that the more confusable with a given

predictor	estimate	SE	z value	
intercept	-3.06352	0.25641	-11.948	***
frequency difference	0.71426	0.22528	3.171	**
similarity index	0.03778	0.08790	0.430	
neighbourhood density	-0.00546	0.21463	-0.025	

Table 5.10.: Summary of the model predicting the recognition of the target's strongest competitor from its similarity to the target, their frequency difference, and the target's neighbourhood density
Significance codes: *** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1
Formula: `comp.perc ~freq.diff + simix + ND + (freq.diff + simix + ND|subjID) + (1|onset.targ/stimulus)`

target a competitor has proven in acoustic studies, the more likely it is to be chosen for recognition; and the denser a target's neighbourhood is, the more the responses distribute over the various candidates and the less responses are apportioned to the main competitor. There is a trend for the influence of similarity to the target, but it is very small and the standard error is relatively high. A possible explanation for this lack of effect is that the operationalisation of the similarity does not match the perceptual reality. One shortcoming of the similarity index employed here is that the studies used tested perception in different languages (English [n = 4], German [n = 3][21], and Spanish [n = 1]) that have phonetically distinct realisations of the phonemes and weigh the cues to phoneme identity differently. A more accurate approach would have been to measure the distance between the two phonemes (target and percept) acoustically. However, this procedure is beyond the scope of the present thesis.

Concerning the lack of effect for neighbourhood density, it is possible that the operationalisation of neighbours in terms of phonological features does not capture the perceptual reality well enough ei-

[21] It has to be noted that one of the confusion matrices for German—Jürgens et al. (2007)—differed greatly from all other confusion matrices and the confusions shown do not seem intuitive.

ther (see above). It seems very unlikely, however, that the neighbourhood numbers calculated on the basis of phonological features are so far off the recognition process as to explain the total absence of even a trend of neighbourhood influence on the recognition of the most prominent competitor. Moreover, there was an effect of neighbourhood frequency, based on those same phonological features, in the main regression model presented in Table 5.4 on p. 109, so the neighbourhood clearly plays a role. Likewise, recent research findings suggest that simple neighbourhood density does not capture the complex processes and dependencies observed concerning neighbourhood influences. For example, lexical neighbours of a target word which themselves have many neighbours show attenuated effects on target word processing, as the target's neighbour also suffers from competition with its own neighbours (Vitevitch & Luce, 2016).

Nevertheless, the significant neighbourhood frequency effect in the main regression analysis again suggests that these simple measures are fine enough to capture the competition processes during perception fairly well. In the main analysis even simple neighbourhood density, which was also tested in a separate model in exchange for accumulated neighbourhood frequency, showed an effect similar to that of neighbourhood frequency. Hence, simple neighbourhood measures such as neighbourhood density and accumulated neighbourhood frequency may be viewed as a rough approximation to a complex field. Even though they might not fully capture the fine inter-dependencies of neighbourhoods, they do show strong and reliable effects, suggesting that the simplification is adequate. Moreover, the more complex neighbourhood measure of relative neighbourhood frequency (i.e., relative to target cluster frequency) has been tested in a separate model and did not yield stronger effects. It remains to be seen whether more sophisticated measures from network science, such as "neighbours of neighbours" (Goldstein & Vitevitch, 2014), lead to even more accurate results. Yet, the discrepancy in neighbourhood effects in the main analysis and the analysis of main competitor choice remains startling. Even though a target's neighbourhood density and frequency do influence whether it is reported correctly, they cannot predict whether the main competitor is falsely perceived instead. It is possible that the percentage of main competitor choices is simply too small (11.0%) to show effects of the coarse neighbourhood measure, while the percentage of correctly reported targets (72.5%) was high enough to allow for modelling with a relatively coarse neighbourhood measure. A more adequate approach to modelling the activation-related processes during perception would be to not limit the investigation to the main competitor but take the similarity and frequency of all neighbours into account. Again, this is not possible for the present thesis.

It would be desirable for future research, however, to model the data from the recognition experiment in PARSYN and ART and compare the results to get a better idea of the competition processes. For now, it can only be stated with some degree of certainty that the strongest competitors were chosen for recognition partly because of their higher frequency as compared to the targets.

5.6.4. Sonority

The hypotheses regarding sonority derived from phonological theory were that consonant clusters conforming to the SSP would be perceived better than clusters violating it and that misperceptions would tend to improve the sonority profile of a cluster. Neither is supported by the data. Instead, an "anti-sonority effect" emerged, that is, clusters violating the SSP have a lower error probability than the ones conforming to it (see Figure 5.2d on p. 110). Moreover, in the few cases of perceptual permutation of consonants, the clusters were almost always "repaired" against sonority sequencing (compare also with Newton's 1972 data and the production study in this thesis, where this phenomenon is much more frequent). This is unexpected in light of phonological theories featuring sonority as a linguistic principle relevant to both language structure and change and psycholinguistic processes such as language acquisition.

The results concerning sonority also deviate substantially from those of a consonant cluster rating study by Albright (2007b). While both (generalised) statistical learning and prior sonority biases make separate contributions to Albright's model, only statistical learning shows the expected effect in the present model. Sonority even seems to impede perception. The most obvious difference between the studies (besides the task difference—listening vs. rating) is that only attested consonant clusters (in some cases marginally attested but nonetheless legal) were used in the present study, while Albright attempted to develop a model that accounts for well-formedness judgements of both attested and unattested consonant clusters. An intuitive explanation for the diverging results is therefore that prior biases can be overridden by statistical learning and are thus only visible where it does not apply. However, inspecting his set of test clusters more closely, reveals that Albright (2007b; p. 4) only used stop-initial clusters and thus did not run into conflicts with sibilant-initial clusters. Conversely, in an fMRI study involving auditory and visual presentation of pseudowords that did include sibilant-initial clusters, Deschamps et al. (2015) did not find an effect of sonority in the auditory modality and conclude that, "while sonority is an important concept in phonological theory, language acquisition and language breakdown", its effect on neuronal activity during phonological processing is limited and absent in auditory processing (p. 82). Hence, the presence or absence of a sonority effect might well depend on whether sibilant–stop clusters are included in the test set or not.

In an attempt to settle the matter, Wright (2001; p. 271) suggests a revision of the SSP, proposing that segments within a syllable should be ordered according to cue robustness (reversing the places for fricatives and stops in the hierarchy) instead of stricture: "segments should be ordered so that transitions from one into the next provide sufficient information for the lexical item to be recovered under normal [i.e. noisy] conditions". This reordering is consistent with the finding in this study that fricative–stop clusters are more resistant to misperceptions than stop–fricative clusters. Whether this imbalance is due to differences in cue robustness or in frequency is still unclear as the two coincide in most cases[22], and it is reasonable that phonotactic structures which have robust cues and are therefore at a perceptual advantage should survive phonological changes better than structures with weak cues. The only exception is the affricate /t͡s/ which is highly frequent in German but exhibits weak robustness concerning cues for the /t/ component. Its low error probability in the experiment data—especially in comparison to /ks/ and /ps/—suggests an influence of frequency on perception, which compensates for the acoustic disadvantage.

Generally, a number of researchers have presented accounts that link differences in the perception of different phonemes to two factors: cue precision and cue robustness (e.g., Baroni, 2014; Henke et al., 2012). Cue robustness, as explained earlier, denotes the perceptibility of a segment's identifying cues under normal listening conditions, and cue precision refers to the degree to which the cue distinguishes the segment from its competitors in perception (cf. Henke et al., 2012). This approach is intuitively more suitable for application to perception phenomena, and it has been shown in Section 5.6.1 that acoustic cues can also explain a great deal of the variation in the data from the present experiment. It is not completely unrelated to the sonority approach, either. The fact that the sonority scale underlying the SSP "corresponds roughly to relative intensity and relative duration" (Henke et al., 2012; p. 97) explains its success in a number of phenomena, but "these dimensions also factor into robustness and recoverability, and therefore do not distinguish a sonority scale from a perceptual scale." Indeed,

[22] It is exactly the frequent violations against the SSP by fricative–stop clusters in many languages that motivated Wright to propose the revision.

Henke and collaborators demonstrate that a phoneme hierarchy based on cue robustness and precision makes very similar predictions to the SSP, but it outperforms it concerning the problematic cases of obstruent sequences. Also in the data from the present experiment, an approach based on acoustic cues (recall the cue-based cluster classification) made better predictions concerning perceptibility of clusters than sonority sequencing. So there is no need to resort to sonority; a cue-based account makes better predictions both for the perceptibility of sequences and—ultimately related—for universal sequencing preferences.

It might be asked whether the binary sonority measure distinguishing only between SSP-adhering and SSP-violating clusters is too coarse to show fine-grained, meaningful sonority effects. To test for this possibility, the model with sonority distance as a predictor instead of SSP violation was set up. However, that model did not converge, and sonority distance only showed a marginally significant effect on one level (Sonority Distance = 1), which is in conflict with sonority-based predictions. Moreover, the finer measure of sonority distance does not solve the problem of sibilant and stop ordering on the hierarchy discussed above. It can therefore be abandoned as an alternative to SSP violation as a predictor both on theoretical grounds and based on model performance. Thus, sonority and the SSP in their present form are untenable as relevant principles for speech perception and should be replaced by sequencing principles based on cue robustness and cue precision. The SSP has been shown to make some valuable cross-linguistic generalisations concerning preferences of linguistic structures, but the cases that have proven problematic there, the sibilant–stop clusters, are the same cases in which the data from the present experiment is in conflict with the hypothesis. All sibilant–stop clusters tested in the perception experiment have very low error rates, whereas their phonological markedness should put them at a perceptual disadvantage if sonority was a relevant principle in speech perception. There are also some clusters whose error rates are notably higher than sonority sequencing would predict. They concern stop-first clusters (followed by a sibilant or liquid), which are harder to perceive for acoustic reasons, as was discussed above. A revised hierarchy as proposed by Wright (2001) would account for both cases. If sonority plays a role in consonant cluster processing (as studies comparing illegal clusters of differing sonority conformity suggest, e.g., Berent et al., 2007; see also 5.2.5), it is probably mediated by perceptual principles like cue robustness and cue precision. In the present study, too many of the test clusters belonged to the problematic obstruent–obstruent cluster groups for which sonority sequencing and cue robustness-based sequencing make diverging predictions. This choice of clusters most likely caused the apparent anti-sonority effect. It can therefore be concluded that a sequencing principle for phonemes has explanatory power only if it is seen from a phonetic–perceptual perspective; as a phonological concept based on abstractions like phonological features it cannot make sufficiently accurate predictions about phonotactic distributions and much less about speech processing.

5.6.5. Salience

The lack of a salience effect in the model is surprising given its foundation in acoustics. It seems that salience as defined by Baroni (2014) and adopted here is an unsuitable operationalisation for the study of perception of sound sequences: whether or not a speech sound is the most salient in its class says nothing about the interplay of cues and modulation of neighbouring segments. Also in Baroni's experiment, salience did not have a significant effect on the perception of obstruent clusters—it could only account for error rates in (plateau) nasal and liquid clusters. The author therefore concedes that the salience scale for obstruents might be wrong. In conclusion, the concept of phonetic salience should be revised to accommodate the findings by Henke et al. (2012) and of the present study.

5.6.6. Summary

Summing up, the present study largely replicated earlier findings on the importance of acoustic cues to speech perception. It expanded on them in showing that in addition to cues, the usage-based measure of sublexical frequency also plays a significant role in perception. In the case of an acoustically ambiguous signal, consonant clusters of high distributional frequency are more likely to be perceptually preserved than those of low distributional frequency. Moreover, they attract responses from their LF neighbours. The phonological neighbourhood of a cluster, in turn, constitutes competition in recognition. The denser it is and the more frequent the neighbours are, the stronger the competition and the lower the recognition rate of the target.

5.6.7. Conclusions and future directions

The findings thus support the view that the bottom-up signal plays the most crucial role in speech recognition but that perception in noisy conditions can be strongly influenced by the distributional properties of the language. The study also calls into question the relevance of phonological principles, namely the Sonority Sequencing Principle, for speech perception. The theoretical notion of sonority may well be a construct that has little to do with the concrete experience of listeners. As it is to some extent correlated with acoustic–phonetic variables, however, it can make correct predictions for speech processing, given a suitable set of test cases.

The findings of the present study have some implications for models of speech perception. First of all, they support interactive activation models featuring activation of several candidates and their subsequent competition for recognition. The results of the study are in line with both PARSYN and ARTPHONE. Two of the three major factors for speech perception in PARSYN—frequency (in this case: sublexical) and neighbourhood—have shown significant effects in the present experiment. The cluster frequency effect found in the experiment can be said to result from the facilitative lateral connections on the pattern level in adjacent positions, and the neighbourhood effect would be considered a consequence of inhibitive connections between pattern level nodes in the same position. The third factor, acoustic similarity to the external input, has not been tested thoroughly. It was entered into the model for percepts, but only as a very rough measure (similarity between the target and its most prominent competitor averaged over confusion matrices from eight perception studies). Taking into account the similarity of the target to several competitors might have produced more meaningful results. There is no doubt that acoustic similarity between the external input and the candidates in the mental lexicon is highly relevant to speech perception.

In ART, the frequency effect would be derived from top-down activation based on learned expectations from the list chunks to the item nodes. So the high error rates of infrequent clusters like /ks/ and /ps/ can be explained by their low expectancy and top-down suppression during the top-down matching process. What makes ART especially appealing for the present study is the fact that the list chunks can take any size, so the frequency of a consonant cluster as such is a relevant entity. Smaller list chunks can indeed be masked by larger ones, but since the pseudowords used in this study did not have lexical frequency values themselves (and in many cases had a syllable frequency of zero), the consonant clusters were the largest units that frequency values were available for. The fact that the same phonemes had varying error rates depending on which cluster they appeared in can be taken as a clue to cluster frequencies overriding those of single phonemes. As in PARSYN, the neighbourhood effect can be explained by the activation of list chunks that share features with the auditory signal. The more expected (i.e. frequent) these competitors are, the more powerful their resonances can be and the more likely it is that one of them reaches a resonant state instead of the target. The unexpected effect of sonority is hard to bring into accordance with either PARSYN or ART, but as discussed above, it is most likely an artefact and not a true effect.

The current results further suggest that only models of speech perception that include either forward and backward transitional probabilities between phonemes or frequencies of larger subsyllabic units such as consonant clusters give a realistic picture of speech perception. Not only did the identification of a /ʃ/ in C1 position influence the identification of the following consonant (towards the high-probability sequences /ʃt/ and /ʃp/), but identification of /s/ in C2 position biased identification of the previous stop even to a greater extent towards /t/ in cases in which /p/ or /k/ had been presented. The fact that later occurring information has an influence on previously occurring information has already been observed in a number of studies (e.g., Cluff & Luce, 1990; Cutler et al., 1987; Ganong, 1980). As mentioned in Section 4.4.1, backward transitional probabilities are implemented in PARSYN via weighted facilitative connections between pattern layer units in successive temporal positions. In ART, such effects are explained by the difference between presentation and recognition rate or more specifically the time the resonance wave takes to develop. So both models can account for the influence of a later-coming segment on a preceding one evidenced in this experiment.

There are some limitations to the study that should be mentioned. Most importantly, it cannot be determined with certainty whether the frequency effect is due to a deliberate response strategy or subconscious mechanisms. A finer measure, such as reaction times, might shed more light on this issue in future studies (cf. Ulbrich et al., 2016).

It should also be kept in mind that with only 16 different onset clusters, predictive power is limited. It is desirable that the study be replicated with more clusters to determine stable effects.

Moreover, the numbers for /sp/ are not completely reliable due to the need to exclude cases where it was written down as <sp> (denoting /ʃp/). This exclusion was made also at the expense of some interesting comparisons which could have provided some insight into the status of acoustically advantageous but infrequent clusters. Moreover, as the discussion showed, it seems promising to set up an alternative segment hierarchy based on cue robustness and cue precision and include it in the analysis to determine the relative contributions of acoustics and the frequency of use. Future studies should pursue this track. However, it is difficult to disentangle their relative influences on perception with certainty since, for obvious reasons, perceptual robustness and frequency of use are usually correlated. It should also be mentioned that only natural speech was used in the experiment as this was deemed important to ensure ecological validity. However, in order to reduce unwanted variability in the stimuli, synthesized stimuli could be used and the results compared to those of the present study so as to disentangle systematic effects due to phoneme-inherent characteristics from the effects of speech variability. Moreover, the same female speaker produced all the stimuli, so there was no way of controlling for speaker-specific effects. It was attempted to minimise these effects by using a trained speaker and recording all stimuli several times and choosing the best tokens, though.

All in all, the present study suggests that consonant clusters are relevant units in speech processing and that the same mechanisms are active on this sublexical level as have been shown to operate on the lexical level. It also showed the significant influence language use has on speech processing. However, it is yet unclear whether the effect of German cluster frequencies are due to a general bias for native-language phonotactics, which the listener has so much experience with, or whether it reflects the specific application of target-language phonotactics to guide listening. In order the disentangle the two, the experiment was repeated with a group of L2 listeners. The results will be reported in the next chapter.

6. Experiment 2: Non-native perception of German consonant clusters

6.1. Introduction

Not only is the way we process speech determined by the phoneme inventory and vocabulary of our native language (Cutler, 2012), but it is also influenced by the structural characteristics of the L1 in almost every way imaginable, including fine phonetic detail (Davidson & Shaw, 2012) and—as was shown in Experiment 1—phonotactics. Previous studies have shown that phoneme sequences which are illegal in the listener's L1 are harder to perceive and are more prone to perceptual illusions, which adjust them to meet the L1's phonotactic rules. For example, word-initial /dl/ and /tl/ clusters are mostly perceived as /gl/ and /kl/, respectively, by native French listeners (Hallé et al., 1998). The effect of perceptual repair of illegal sequences in non-native listening has been repeatedly found (e.g., Dupoux et al., 2001; Hallé et al., 1998; Massaro & Cohen, 1983; Pitt, 1998). However, these studies test L1-illegal phoneme sequences, that is, sequences that are unattested in the listeners' mental lexicon. What happens in the perception of phonotactic structures when a language user is familiar with two differing phonotactic systems? Can this knowledge of an additional system influence the perception of the other language? Comparatively few studies are concerned with how gradient phonotactics of one familiar language affect processing of another. It has been found that L1 knowledge of a less restrictive phonotactic system reduces difficulties in processing and memorising L2-illegal consonant clusters (Ulbrich & Wiese, 2018). In the same vein, knowledge of a less restrictive L2 reduces perceptual repair effects in L1-illegal clusters; this effect is gradual: the higher the proficiency in the less restrictive language, the weaker the perceptual repairs (Carlson et al., 2016). This is not surprising considering the need to broaden phonotactic restrictions in order to more faithfully perceive a less restrictive language. Carlson et al., however, only examine categorical phonotactic effects, namely the attenuation of perceptual illusions for consonant clusters that are illegal in the L1 if they are legal in the L2. Similarly, most studies are concerned with categorical phonotactics, examining the processing of sound sequences that are illegal in the hearers' L1 and/or L2. Less is known about how gradient phonotactics, that is, frequency distributions, of the languages in question influence L2 speech perception. The few studies that examine gradient phonotactics reach conflicting results. Some find only an influence of L1 phonotactics (Lentz, 2011; ch. 3), some only an influence of L2 phonotactics (Lentz, 2011; ch. 4; Lentz & Kager, 2015). Hanulíková et al. (2011) observed effects of both L1 and L2 consonant cluster frequencies, while Cohen et al. (1967) did not find an influence of either.[1] However, these studies investigated speech processing on the lexical level. It is the aim of this chapter to shed light on how knowledge of both L1 and L2 gradient phonotactics influences perception of phoneme sequences and their processing on the sublexical level in the L2. In order to do so, the perception experiment from the previous chapter was repeated with a group of English learners of German.

[1] The main object of investigation in the latter two studies was categorical phonotactic effects, but they consider gradient phonotactics in separate analyses.

6.2. Previous research

6.2.1. Characteristics of L2 listening in noise

It is well-known that noise-masking affects non-native listening more than native listening (e.g., Lecumberri & Cooke, 2006). In L1 phoneme identification, listeners use multiple, redundant cues and cue-weighting to overcome the effects of energetic masking. In L2 listening, on the other hand, fewer cues and less sophisticated weighting strategies are available due to limited exposure and less experience with noisy conditions (Lecumberri & Cooke, 2006). In noise, the richness or paucity of the cues used can be decisive. The cues that L2 listeners usually attend to may be masked, while cues additionally employed by L1 listeners may withstand masking and still be available. L2 listeners are therefore expected to perform worse overall in comparison to the L1 listeners in the previous experiment. Moreover, the kind of cues used to identify a phoneme is also heavily influenced by the L1 (Davidson & Shaw, 2012; Lecumberri & Cooke, 2006). On the other hand, an EEG study found that L2 listeners attend more to acoustic detail than L1 listeners (Song et al., 2019). This might compensate for some of the difficulties, albeit only for phonemes with clearly perceptible cues.

6.2.2. Phonotactics in L2 perception

As mentioned above, the bulk of research on phonotactic effects in L2 perception is concerned with categorical phonotactics. Taken together, these results suggest that L2 listeners are able to use their knowledge regarding the legality of L2 phonotactic sequences for speech processing. For example, Weber and Cutler (2006) examined the performance of highly proficient German learners of English in a word-spotting task. Their results showed that L2 listeners are sensitive to English phonotactic boundary cues (sequences that are illegal word-initially, such as /ʃl/) even when they deviate from the rules in their L1; L2 listeners can make use of such cues in speech segmentation almost as well as L1 listeners. However, L2 listeners were also influenced by German boundary cues that do not apply to English phonotactics (sequences that are illegal initially in German but not English, such as /sl/), which are hence not relevant to the L2 listening task. Thus both L1 and L2 categorical phonotactics exert an influence on word segmentation.

Similar results have been obtained by Hanulíková et al. (2011): they studied whether native speakers of Slovak, who have been shown to be unaffected by the Possible Word Constraint (PWC) when segmenting their L1 (a language which allows single consonants as words; Hanulíková et al., 2010), use the same strategy when segmenting German (a language in which the PWC is in force). Their results indicate that Slovak learners of German are aware of the phonotactic differences between their L1 and L2 and show effects of the PWC when segmenting German. However, even though they clearly differed from the group instructed to segment Slovak nonce words, they also showed slight effects of Slovak phonotactics (more specifically, they segmented nonce words faster in contexts where a consonant constituted a Slovak word in comparison to when it did not constitute a Slovak word). This means that, for segmentation of the L2, the influence of L2 phonotactics was strongest and clearly aided speech segmentation, but there was also a small effect of L1 phonotactics which, although much weaker, was the same in nature as that observed during L1 segmentation.

Likewise, Carlson (2018), Carlson et al. (2016), and Ulbrich and Wiese (2018) found effects of both L1 and L2 phonotactic systems in L2 perception. Carlson et al. (2016) investigated the question of how bilinguals perceive consonant clusters that are legal in one of their languages and illegal in the other. They had Spanish–English bilinguals and monolingual controls execute a vowel recognition task and an AX discrimination task with stimuli consisting of acoustically reduced vowels (/a/ or /e/) followed by sC clusters. They found stronger effects of perceptual repair for Spanish monolinguals and Spanish-dominant bilinguals. English-dominant subjects showed the weakest effects of illusory vowels, which clearly demonstrates how relative language proficiency and dominance modulate the strength of the phonotactic effect. As mentioned above, they interpret these results as indicating that knowledge of a less restrictive phonotactic system can reduce perceptual repair effects, even when the more restrictive language is dominant. However, the alleviating effect of the less restrictive language becomes stronger with increasing language proficiency. The authors conclude that knowledge of two phonotactic systems can be integrated and jointly influence speech perception. In a follow-up study, Carlson (2018) explored whether this reduced illusory effect in bilinguals is due to an added representation for a phonotactic sequence (namely the unaltered consonant cluster, which stems from the L2) competing with the repaired representation of the L1 or whether it is due to a retuning of perception in compliance with the L2. Using the same kind of stimuli as in Carlson et al. (2016), but with the added manipulation of language mode in his bilingual subjects, he found that perceptual illusion is weakest when his subjects were set into English mode prior to the listening task (i.e., when their English system was co-activated). He concludes that perceptual repair is not weakened per se but that another, competing, representation is added to the bilingual listeners' mental lexicon, which causes a reduction in the effect.

A few studies failed to find effects of either L1 or L2 phonotactics, however. In a lexical decision experiment, Trapman and Kager (2009) found phonotactic L2 knowledge of onset clusters to influence both accuracy scores and reaction times in advanced Spanish and Russian learners of Dutch (as well as Russian beginners), but there was no difference in their performance for consonant clusters that are vs. are not part of their native repertoire. Spanish beginners, in contrast, who had not yet acquired Dutch phonotactics, showed no difference between Dutch-legal and Dutch-illegal clusters. However, they did not differentiate between Spanish-legal and Spanish-illegal clusters, either. Thus only L2 phonotactic knowledge affected learner performance, and in the case of a phonotactically more restrictive L1 (Spanish), it only affected advanced learners.

Kabak and Idsardi (2007), on the other hand, examined whether L2 misperceptions of (ambisyllabic) consonant clusters that are illegal in the L1 are due to syllable structure restrictions (i.e., abstract phonotactic patterns) or the legality of concrete consonant sequences in the L1. They had Korean subjects solve an AX discrimination task with English nonce words that contained either coda consonants or coda-onset sequences which are illegal in Korean. Since participants only had problems perceiving (Korean-) illegal coda consonants and not illegal coda-onset groupings, the authors conclude that it is syllable structure restrictions in the L1 that cause misperceptions in L2 listening. Consequently, there is an effect of L1 phonotactics, but it consists of abstract phonotactic rules rather than phonotactic sequences.

Taken together, studies that investigated the effects of categorical phonotactics in L2 perception have found that there is ample influence from both L1 and L2 phonotactics. Most studies reviewed here found simultaneous effects of both. Kabak and Idsardi (2007) only tested for L1 phonotactic influence and found evidence solely for effects of syllable structure rules, not for the legality of individual consonant sequences. Trapman and Kager (2009) found only L2 phonotactics to be relevant.

More importantly, however, there are a few studies that take gradient phonotactics into consideration. These suggest that L2 learners not only make use of illegality knowledge in the L2 but also L2 frequencies during different speech processing tasks. For example, high L2 biphone frequencies facilitate learning of syllable structures (especially those that are illegal in the L1), as well as later recognition (Boll-Avetisyan, 2012). In their study on the PWC, Hanulíková et al. (2011) assessed effects of consonant cluster frequencies. They found that Slovak learners of German are to a larger degree influenced by Slovak frequencies of initial consonant clusters when segmenting German speech than by the German frequencies.[2] The assumption that learners are influenced by L1 frequencies rather than the frequencies of the L2 they are listening to is very plausible in light of usage-based linguistics since L2 learners have much more experience with the structure of their native language.

Converging evidence comes from Lentz (2011), who also examined whether and how L2 listeners are influenced by L2 gradient phonotactics during segmentation. He conducted an eye-tracking study based on the visual world paradigm (Tanenhaus et al., 1995) but presented four letters (representing potential word starts) instead of four pictures during each trial. The direction of subjects' gaze to the individual letters was taken as an indication of the segmentation hypotheses at each point in time. In the experiment, the gaze of L1-Slavic learners of Dutch revealed that they had a tendency to hypothesise word boundaries as falling between the two consonants of a cluster that is illegal in Dutch. This result parallels the behaviour of Dutch L1 listeners obtained in an earlier experiment (Lentz, 2011; p. 125). However, in trials where there was no word present, their gaze indicated that they also considered segmentations in which a Dutch-illegal consonant cluster would constitute a word onset—in contrast to L1 listeners' behaviour. This led Lentz to conjecture that L2 listeners do not acquire L2 legality but are instead led by the gradient phonotactics of their L1, which ascribes a lower degree of well-formedness to the Dutch-illegal as compared to the Dutch-legal clusters. This low degree of well-formedness seems sufficient for a segmentation hypothesis in cases where there is no segmentation of higher phonotactic well-formedness leading to viable lexical activation. Hence the results of this experiment suggest an

[2] At the same time, participants were able to apply rule-based segmentation strategies that are specific to German but do not hold for Slovak speech segmentation—namely application of the PWC—which shows that they are proficient L2 users and, generally, they are able to make use of properties of the L2 phonological system. This suggests that application of gradient phonotactics is more difficult than utilising absolute principles and is acquired late in the L2 acquisition process.

influence of L1 gradient phonotactics, although it is not possible to distinguish between the influence of gradient L1 phonotactics and (partly) acquired categorical L2 phonotactics with certainty. It is important to note that the greatest difference between L1 and L2 listener behaviour was observed in the nonword trials. In trials in which target words were present, the effects of phonotactics were most likely attenuated by stronger lexical effects in the recognition process. This means that using only nonce words in the present study is promising in terms of visible effects of phonotactics also in L2 listeners.

Finally, Lentz and Kager (2015) tested for both gradient effects of L2 phonotactics and categorical effects of L1 phonotactics and found both to be present simultaneously: L2 gradient phonotactics can be acquired but this is done via an L1 phonotactic filter, which causes (perceptual) vowel epenthesis in L1-illegal consonant clusters. In a lexical decision task with priming, their subjects—Spanish and Japanese learners of Dutch—showed that they had learned which Dutch consonant clusters are frequent and which ones are infrequent (although still legal) and were able to use this knowledge during speech processing. Since they were also primed by epenthesised clusters, the authors concluded that L2 learners acquire this gradient L2 phonotactic knowledge through an L1 filter; that means their representations of frequent and less frequent Dutch consonant clusters were not faithful but were corrected according to the phonotactic rules of their respective L1s. Conversely, the results could also be interpreted in accordance with Carlson's (2018) explanation that bilinguals have two competing representations, one from each language. In this case, too, the epenthesised prime would prime the non-epenthesised target.

The studies summarised so far suggest that, while L2 listeners are able to make use of L2 statistical knowledge for speech processing, they are nonetheless also influenced by the structural properties of their L1, even when it is not of any assistance in L2 listening. Learners cannot completely inhibit irrelevant L1 knowledge during L2 perception. This raises the question of whether they have separate phonotactic systems for their languages or whether they have acquired "phonotactics" as a whole—aggregated over the various input languages. Hanulíková et al. (2011; p. 516) also note that, "To a large extent, learners' frequencies might be determined by the L1 as well as by a subset of L2 (most likely the more frequent structures [...])". Two factors that are potentially crucial to the relative influence of the L1 and L2—language mode and proficiency—will be discussed in the next section.

It is also important to note that both Hanulíková et al. (2011) and Lentz (2011) investigated the influence of cluster frequencies on segmentation processes. It is possible that listeners draw on different resources for tasks that do not involve segmentation of speech.

Language mode

Language modes in bilinguals, a concept described in detail by François Grosjean (e.g., Grosjean, 2001), stretch from a totally monolingual mode at one end of the continuum to a bilingual mode with a high degree of mixing at the other end. Language mode has been shown to influence different aspects of speech processing in bilinguals (e.g., phoneme perception: Elman et al., 1977; lexical access: Dunn & Fox Tree, 2014). For the sake of simplicity, the whole spectrum of monolingual to bilingual language modes will not be considered here; instead, only L1 and L2 mode will be differentiated. Some results suggest that, depending on the language mode in an experiment, participants exhibit different kinds of behaviour, also with respect to phonotactics. A non-target language that is highly activated will interfere strongly with the phonotactics of the target language, as Freeman et al. (2016) demonstrated. In an L2 lexical decision task with cross-modal priming, in which L1 activation was deliberately induced through the use of cognates among the primes, they found evidence of activation of non-target language phonotactic constraints: Spanish-English bilinguals were primed for

/st/ and /sp/ onset targets (e.g., *stable*) by primes beginning with an /ɛ/ onset (e.g., *elopevent*). This can only have been caused by activation of the Spanish epenthesis (prothesis) rule, which is active for /st/ and /sp/ onsets. Crucially, the epenthetic effect also occurred in non-cognate primes, such as *strong* (Spanish: *fuerte*), which cannot have been caused by lexical activation of the cognate. This demonstrates an effect of activation of non-target language phonotactics. It would be insightful to repeat the experiment without cognates in the set of primes in order to explore whether non-target language phonotactics are also activated when overall L1 activation is smaller. It can be assumed that the L1 phonotactic effect, if at all present, would be smaller in such cases.

Support for the assumption that this kind of manipulation causes a language-mode effect comes from Lentz (2011), who repeated two versions of his visual world paradigm experiment: an English language version and a Dutch language version. In the English language version (in which instructions and target words were in English), the Dutch participants displayed more fixations on English-illegal/Dutch-legal clusters than in the Dutch language version. This shows that they are able to switch between their phonotactic systems and use the appropriate one for segmentation depending on the situation. Lentz therefore concludes that "the representation [of phonotactic knowledge] is labelled as belonging to one of the two languages." (Lentz, 2011; p. 168) and, in cases where the two phonotactic systems make conflicting predictions concerning segmentation, the language mode of the listener determines which one is employed. This interpretation contrasts with Weber and Cutler's (2006) and Hanulíková et al.'s (2011) results that, even when segmenting L2 speech, learners are influenced by L1 phonotactics. What causes these diverging results? Although not explicitly mentioned in the paper, it has to be assumed that the experiment by Weber and Cutler (2006), too, was designed to force the subjects into L2 mode (i.e., used instructions in English etc.). In the study by Hanulíková et al. (2011), they also attempted to induce L2 mode by giving participants instructions in the L2. As the authors note, however, the subjects were aware of the required L1 proficiency and many of them lived in an L1 environment, hence the language mode may have been ambiguous. The same can be said of the subjects in Lentz's study, however. It is therefore unlikely that differences in language mode caused the conflicting findings. The L2 proficiency of the participants across the studies was comparable (although the professional interpreters participating in Weber and Cutler's study were probably slightly more proficient in their L2 than the participants in the other two studies). One notable difference between the experiments is that Lentz (2011) explicitly excluded lexical effects by only analysing data from nonword trials, whereas Weber and Cutler (2006) analysed only word trials, and Hanulíková et al. (2011) used word targets in all critical trials, leaving room for lexical effects. If this is indeed the source of the different behaviour, the results in the present study can be expected to resemble those of Lentz (2011) because, like in his experiment, the potential for interference from lexical effects has been minimised.

L2 Proficiency

L2 proficiency plays a role in the exploitation of L2 categorical phonotactic knowledge, both for the attenuation of perceptual illusions and for word recognition. The study by Carlson et al. (2016) detailed above shows that the occurrence of perceptual illusions in bilinguals in sequences that are legal in one language but illegal in the other is modulated by language dominance and the level of L2 proficiency. Moreover, the effect of L2 legality on word and nonword recognition increases as a function of L2 proficiency (Trapman & Kager, 2009).

Ulbrich and Wiese (2018) note that the level of proficiency in the L2 influences to what extent L1 phonotactics affects L2 processing. Based on their experiment—a CVCC-nonword learning paradigm for Chinese and Russian learners of German—they assume an L1 phonological filter through which the L2 is perceived. This filter is reduced with increasing L2 mastery. Crucially, they also postulate an L2 phonological filter whose impact depends on its constraining strength relative to that of the competing L1 filter: The more rigid the constraints imposed on phonotactic sequencing by the L2 filter (compared to those of the L1), the stronger they are enforced (Ulbrich, personal communication). However, L2 constraints can "shape language-specific phonotactic mechanisms in both directions" (Ulbrich & Wiese, 2018; p. 178), that is, not only expand them as they did for their Chinese subjects, but also limit them, as they did for their Russian subjects. Since an L2 filter is acquired by learners, its strength also depends on L2 proficiency of course. For this reason, Ulbrich and Wiese's advanced Chinese subjects were better at recollecting nonce words with existing German consonant clusters, while their Chinese beginners were not. These data suggest that the extent to which L2 learners rely on L1 vs. L2 phonotactics in their speech perception strategies is a function of their L2 proficiency, with more proficient L2 users being able to ignore native phonotactics and utilise L2 phonotactics instead (see also Lentz & Kager, 2015; Weber, 2001; for similar stances).

Note, however, that Altenberg (2005) did not find an interaction between L2 proficiency and L2 legality on the perception of consonant clusters.

In terms of gradient phonotactics, proficiency has been found to modulate the effect of sublexical frequencies on nonword recall and nonword perception. In a short-term memory nonword recognition task, recollection of nonwords of high phonotactic probability improved with increasing L2 proficiency, but, for nonwords of low phonotactic probability, L2 proficiency was not a relevant factor (Boll-Avetisyan, 2012). In a primed lexical decision task for Japanese and Spanish learners of Dutch, Lentz (2011) found that facilitation of nonwords with HF consonant clusters (as opposed to medium-frequency clusters) became stronger with increasing L2 proficiency for L1 Spanish learners. At the same time, this took place through an L1 phonotactic filter (see above). For the Japanese learners, there was no facilitation of HF clusters and no proficiency difference. The results are therefore mixed.

Summary

Summing up, there is ample evidence for influence of both L1 and L2 phonotactics on L2 perception. Most studies found effects of categorical phonotactics, that is, the legality of sequences. Knowing a less restrictive language can reduce perceptual illusions, which occur in sequences that are illegal in the listener's phonotactic system. On the other hand, learning a more restrictive system can lead to subsequent differentiation between well-formed and ill-formed phonotactic structures during language processing. Legality of L2 structures thus seems to be easily acquired and used in subsequent L2 processing. It is less clear if gradient L2 phonotactics can be acquired as quickly as categorical phonotactics. Results by Hanulíková et al. (2011) suggest that in terms of gradient phonotactics, L1 influence is stronger than L2 influence during L2 segmentation, even for proficient L2 users. To my knowledge, no studies thus far have directly compared the effects of L1 and L2 gradient phonotactics on processing accuracy, but influences of both have been found in separate investigations (L1: Lentz, 2011; L2: Boll-Avetisyan, 2012; Lentz & Kager, 2015; Trapman & Kager, 2009).

Crucial influencing factors for the relative influence of the two phonotactic systems are language proficiency and dominance, as well as language mode, during and prior to the perception task.

As in the L1 listening experiment, the facilitation of certain consonant clusters will not only be investigated in terms of language-specific phonotactics but also with regard to universal sequencing preferences based on sonority. The next section will summarise the previous literature on sonority sequencing effects in L2 perception.

6.2.3. Sonority in L2 perception

Considering the results from Experiment 1, it seems unlikely that sonority differences between the clusters used have an effect on the perception task at hand. It is conceivable, however, that universal principles, such as sonority, are more relevant to L2 perception than L1 perception.

Previous research gives only very limited insight into this question: although studies that tested the effects of sonority sequencing on the perception of L1-illegal phoneme sequences abound (e.g., Berent et al., 2007; Tamási & Berent, 2015; Zhao & Berent, 2016), there are hardly any studies that examine the interplay between sonority and L2 phonotactics. Relevant research in the domain of L2 acquisition is the aforementioned study by Ulbrich and Wiese (2018), who showed that the SSP-conformity of consonant clusters has an effect on nonce word recall both for learners with a phonotactically less restrictive and a phonotactically more restrictive L1. Both Chinese and the Russian learners of German were faster and more accurate in recognising nonce words with coda clusters that conform to the SSP as opposed to those that violate it. Moreover, the effect of sonority was stronger than the effect of L2-legality: this was only relevant for SSP-violating clusters with Chinese learners and led to inconsistent outcomes.[3]

Although they do not investigate sonority directly, some conclusions as to the role of sonority can also be drawn from the results of Trapman and Kager (2009) because the consonant clusters in their study were divided into three groups, which—although these clusters were classified based on legality in the L2, as well as the respective L1s of the partici-

[3] Responses were faster and partly more accurate to stimuli with SSP-conforming but German-illegal clusters.

pants[4]—show some sonority differences: Type 3 clusters (e.g., /fl/, /pr/) show sonority distances of 2–3 and can therefore be considered to be almost optimal, whereas type 2 clusters generally show smaller sonority differences (mostly 1–2, e.g., /sm/) and include the SSP-violating cluster /st/. Type 1 clusters are, on average, the least well-formed and cover a sonority distance range of −3 to 2 (e.g., /rt/, /zl/). In wordlikeness ratings, advanced Russian learners of Dutch differentiated between type 2 and type 3 clusters, both of which are legal in the L1 as well as the L2 but have different levels of SSP-conformity. This means that the advanced Russians, who were familiar with the clusters from their L1 and were probably also aware that they exist in the L2, too, still considered the sonority-wise more well-formed clusters to be better than the less well-formed clusters. Beginner-level Russian learners, on the other hand, did not distinguish between the two kinds, which makes sonority as potentially universal phonotactic knowledge a less convincing explanation. Moreover, even the advanced Russians only displayed such sensitivity in the wordlikeness rating task, not in the lexical decision task, which enables direct insight into speech processing and is therefore more relevant to the present study. It can therefore be assumed that it is the metalinguistic character of the wordlikeness ratings that caused, or at least enhanced, the effect.

In summary, there seems to be some sensitivity to sonority in the processing of L2 sequences, but it does not surface during all speech processing tasks and in all measured variables. Generally, sonority sequencing seems to be more influential in meta-linguistic and recall tasks. In lexical decision, Trapman and Kager (2009) did not find an effect of sonority in L2 learners from different language backgrounds.

[4] Type 1 clusters are legal only in Russian, type 2 clusters in Russian and Dutch (L2), and type 3 in Russian, Dutch, and Spanish.

6.3. English and German phonotactics

Before turning to the L2 experiment itself, a brief comparison of the listeners' L1 (English) and L2 (German) phonotactic systems will be given to serve as a foundation for the interpretation of the experimental results and any potentially language-specific effects. The two languages differ in both categorical and gradient phonotactics.

Since German and English are closely related, however, they are similar in their phonotactic requirements. Both languages allow a number of syllable-initial CC clusters (e.g., German /plat/ "flat", /ʃneː/ "snow"; English /pleɪt/ "plate", /snəʊ/ "snow"). In addition, initial CCC clusters are only licenced if the first consonant is a sibilant—/s/ in English and /ʃ/ or (less frequently) /s/ in German (e.g., German /ʃpʁaːxə/ "language", English /stɹɒŋ/ "strong"). Many initial consonant clusters are the same in the two languages (e.g., /pl/, /tr/, /sk/, /kl/ etc.) or have close parallels (e.g., German /ʃt/, /ʃp/, /ʃl/, /ʃm/, /ʃn/ historically correspond to English /st/, /sp/, /sl/, /sm/, /sn/, as can be seen in a number of cognates, such as /ʃlaːf/–/sliːp/, /ʃtaɪ̯n/–/stəʊn/).

The most notable difference between the two languages is that German allows for initial stop–sibilant (in loans: /ˈpsyːçə/, /ksenofoˈbiː/) and stop–nasal (/pnɔɪˈmaːtɪk/, /gnoːm/) clusters, while English does not (*/ps/, */ks/, */gn/, */pn/). Cognates of words starting with these clusters in German are simply reduced to C2 in English (e.g., /saɪki/, /zɛnəˈfəʊbɪə/, /n(j)ʊˈmætɪks/, /nəʊm/). Moreover, English allows C + glide clusters (e.g., /njuː/, /kwiːn/), which German does not. These are the only areas in which the phonotactics of the two Germanic languages differ with respect to the consonant cluster classes they license.

Furthermore, /t͡s/—the only onset in our test set that is not a true consonant cluster—is not part of the native English phoneme inventory. Nor is there a true consonant cluster composed of /t/ and /s/ in syllable-initial position in English since, as stated above, sequences of a stop followed by /s/ are illegal.[5] Instead, the English phoneme inventory features the affricate /t͡ʃ/, which in turn does not exist in German. Note, however, that although /t͡ʃ/ as an affricate is absent from the German phoneme inventory, the *consonant cluster* /tʃ/ does exist in German words of foreign origin (e.g., /tʃʏs/, /ˈtʃɛlo/) and is included in the test set.

The differences in terms of gradient phonotactics are larger between the two languages. Figure 6.1 displays the English and German log frequencies of the initial consonant clusters used in this study. As the graph shows, some of the native German clusters—such as /ʃt/, /ʃl/, and /ks/—have zero frequency[6] in English, while others (mostly /ʃ/ + sonorant) are not native English clusters; they have frequencies above zero in the CELEX database due to their occurrence in loan words (e.g., /ʃp/ in *spiel*, /ʃn/ in *schnapps*; surprisingly, CELEX also lists two lexemes with initial /ps/: /ˈpsjuːdəʊ/, /psɒˈrøəsɪs/—both of which are commonly pronounced with a simple onset[7]). A third group of clusters has a higher frequency in English than in German and some of them (/sk/, /sl/, /sp/) are native to English but are marginal or restricted to loan words in German. Hence the clusters used in the present study are not distributed evenly along the English frequency scale but can rather be divided into two groups with relatively small frequency differences within the groups: English-HF clusters and illegal (or non-native) clusters.

It should be noted, however, that in spite of having zero frequency according to CELEX (and this is true for Standard English in general; Received Pronunciation as well as Standard American English), [ʃt] has

[5] It does occur, however, in a very small number of loanwords, like /ˈtsɛtsi/, /ˈtswɑːnə/, /tsuːˈnɑːmi/, although these have pronunciation variants without /ts/ as well.

[6] Frequencies given here are log-transformed. Since the logarithm of 0 cannot be computed, 1 was added to all raw frequencies.

[7] König and Gast (2012) note "a recent tendency to pronounce such clusters in accordance with their orthography", which they attribute to hyper-correction or the influence of written on spoken language.

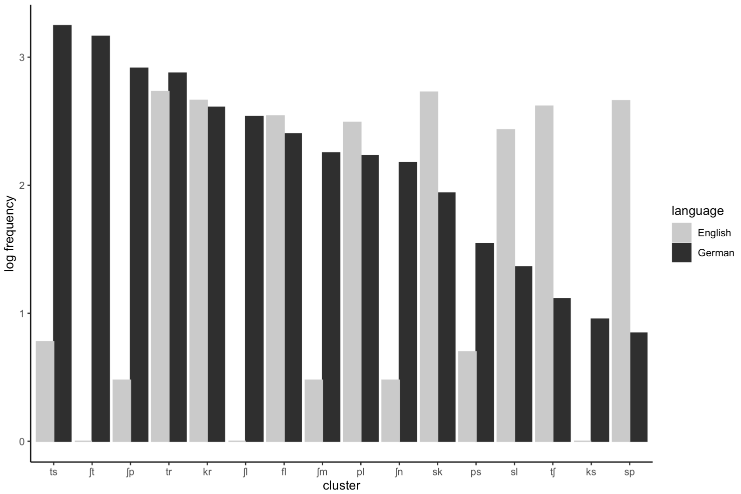


Figure 6.1.: Frequencies of the test consonant clusters in English and German (Clusters are arranged according to their German frequencies from most to least frequent.)

started to occur as a variant of [st] in initial /str/ clusters in some—especially American—English varieties. This is considered a sound change in progress, probably an assimilation triggered by the [ɹ] environment (e.g., Rutter, 2011; Shapiro, 1995; Stevens & Harrington, 2016) but is sporadically spreading to /st/ in other, even r-less, contexts, such as [ʃtɪɫ] (Janda & Joseph, 2003). In Australian English, this change is not currently occurring, but acoustic and perceptual studies show that the sibilant in [stɹ] clusters is retracted, that means, realised in postalveolar position. As a result, it is categorised by native listeners of Australian English as a token of /ʃ/ when spliced into pre-vocalic contexts (Stevens & Loakes, 2019). Thus, despite the phonotactic illegality of /ʃt/ in standard varieties of English, native listeners of Australian English may have some experience with the sound sequence [ʃt], albeit mostly in r-contexts.

6.4. Research questions and hypotheses

Having established in Experiment 1 that L1 listeners use gradient phonotactic information in (nonce[8]) word recognition, the present chapter investigates whether L2 listeners are also able to make use of target language phonotactic distributions. Moreover, the extent to which non-native listeners are influenced by the phonotactics of their L1 during L2 listening will be examined. This is related to the question of whether bilinguals have separate phonotactic systems for their languages, which they can draw upon to support speech processing in the respective situation.

Furthermore, whether or not sonority sequencing plays a role in L2 perception will be investigated. Like in the L1 listening study, the relative influences of language-specific phonotactics (this time, of both L1 and L2) and that of sonority will be compared.

Based on previous research, it is hypothesised that L2 listeners will be able to make use of German cluster frequencies in much the same way as the L1 listeners in the previous experiment did, but that they might be influenced by their native cluster frequencies at the same time, albeit to a much smaller degree. The role of sonority in L2 perception is still an open question. Judging from the L1 perception results (Chapter 5), sonority sequencing is unlikely to show facilitative effects on accuracy rates in nonce word identification in noise. However, L2 listeners, who are not as familiar with language-specific phonotactic distributions in the L2 yet, might rely more on universal principles, like sonority sequencing, than L1 listeners do. Previous studies have come

[8] It might be argued that there is no clear distinction between L1 and L2 perception for nonce words. The nonce word stimuli used in the two experiments are considered to be "German" because they follow German phonotactics (to varying degrees, as indicated by the log frequencies of the test clusters) and were produced by a native German speaker so that their pronunciation follows German phonetics. Furthermore, subjects were instructed to use a system based on German spelling to write down what they heard.

to diverging conclusions, but there are some indications of sonority effects in L2 perception.

6.5. Methods

6.5.1. Participants

Twenty-two learners of German (14 female; mean age: 34.32, SD = 13.67) participated in the experiment and received monetary compensation for their participation. The participants were native speakers of Australian English living in Australia. However, three of the subjects actually acquired German as their first language, but moved to Australia at an early age (6 years, 1 year, exact number of years unknown for the third subject) and regarded English as their native language. They spoke German with a noticeable English accent and made grammatical mistakes. Like all of the other subjects, they took the lexTALE lexical decision test in German (see below, Section 6.5.3). In addition, two of them also took the English version as a point of comparison. Their results for the English version were consistently higher than for the German version (s17: German: 75%; s21: German: 61.25%, English: 97.5%; s22: German: 73.75%, English: 98.75%). One subject reported acquisition of Kannada as his first language. He, too, considered English as his primary language and reported that he only spoke Kannada with his grandparents and at a lower proficiency than English. All participants reported normal hearing. Their self-reported German levels range from B1 to C2 (B1: n=2; B2: n=10; C1: n=7; C2: n=3). Table 6.1 gives an overview of participants' L2 proficiencies according to different criteria.

A few of the L2 level measures in Table 6.1 are correlated. For example, times spent doing different types of activities in German is highly correlated (Pearson's r range between .81 and .98, with $p < .001$). None of the measures are strongly linked to the other measures, though. Self-

measure	mean	sd
length of study (years)	14.11	14.58
age of acquisition	16.41	11.34
duration of stay in German-speaking country (months)	24.06	25.95
time spent on activities in German (times/week)		
reading	27.89	63.93
writing	17.67	59.11
listening	39.17	117.90
speaking	25.14	60.20
lexTALE score (in %)	68.98	8.48

Table 6.1.: Subjects' L2 levels

reported CEFR level correlates most strongly with lexTALE score (Pearson's r = .68, p < .001) and duration of stay in a German-speaking country (Pearson's r = .64, < .001). Otherwise, no two measures have a correlation coefficient above .50, which would be considered the starting point for moderate correlation (Taylor, 1990).

All participants gave informed consent.

6.5.2. Materials

Stimuli

The stimuli used in the experiment were the same as those used in the L1 perception experiment (Experiment 1, Chapter 5).

Multi-talker babble

The multi-talker babble files used in the experiment were the same as in Experiment 1. Again, the babble files were randomly assigned to the stimuli for each subject.

6.5.3. Design and procedure

The main experiment followed the same procedure as in Experiment 1. Twelve of the participants took the experiment in a sound-attenuated booth, while the remaining ten took it in a quiet library room.[9] Before the experiment began, participants were greeted in English (unless they started speaking German on their own accord). The experimenter then told them that she would explain the task in German so that the participant could get into "German mode" and that he/she should ask if anything was unclear or if the instructions were given too fast. The procedure and details regarding the main task were then explained to them in German; that means the maximum amount of English interaction for all subjects was 2–3 sentences. Subjects were told that they would hear nonsense syllables that sound like German words. They received a sheet explaining the spelling system to be used, which was very similar to the one used in the L1 version, but which explicitly referred to some German spelling conventions (e.g., *"ch" wie in "ich"*, "'ch' as in 'ich'"; *"w" wie in "wer" (entspricht engl. "v")*, "'w' as in 'wer' (corresponds to English 'v')") and explicated the spelling difference between <s-p> = [sp] and <sp> = [ʃp] with reference to English and German pronunciations. To increase motivation for faultless transcriptions, participants were told that the five people with the fewest mistakes would go into a raffle for a gift card.

After that, the participants filled in a questionnaire (see Appendix A) on their personal data, language background (including self-assessed German proficiency, age of acquisition, time spent in German-speaking

[9] The results of the hearing screening suggest that the acoustic conditions were comparable in the two settings: In the lab, participants missed seven tones in total (= 3 %) and in the library room, nine tones in total (= 6 %). A logistic regression analysis with with frequency of failure to hear pure tone stimulus as the binary dependent variable, lab condition, frequency of tone, and age of subject as fixed effects and random intercepts for lab condition and frequency of tone by subject did not show a significant effect for lab condition (p > 0.5).

countries and studying German, experience with German dialects and other languages), as well as any potential hearing impairment. Following this, they took the German version of the lexTALE lexical decision test (Lexical Test for Advanced Learners of English, Lemhöfer & Broersma, 2011; www.lextale.com). During the test, 60 German lexemes and nonce words that follow German phonotactics and, in many cases, contain existing German affixes were presented in written form. Participant were instructed to indicate via coloured and labelled buttons on a screen whether the item presented was a German lexeme or not. There was no time limit for the lexical decision task. After the lexTALE test[10], the hearing screening[11] and the experiment proper were carried out. As in Experiment 1, the experimental task was to freely transcribe monosyllabic pseudowords with initial consonant clusters, which were embedded in multi-talker babble.

Like the L1 subjects, the L2 subjects were given ten practice trials with feedback to familiarise them with the task.

6.5.4. Analysis

Data preparation and analyses followed the same modus operandi as the analyses of the L1 data. The dependent variable was again error in the onset cluster (binary variable). Since almost half the /sp/-stimuli were transcribed as <sp>, which seems to suggest spelling rather than perception problems, these cases were similarly discarded. This was

[10] For reasons of time, eight participants took the experiment before the lexTALE test and the questionnaire.

[11] In the hearing screening, one subject missed six of the 15 tones, one missed four, one three, and three missed one each. It turned out afterwards that the participant who missed six tones had her headphones incorrectly connected to the experiment laptop (so that she heard the tones via loudspeaker instead); this mistake was rectified before the main part of the experiment began, so it was decided that this participant's data should be included nonetheless. The other 16 subjects heard all 15 tones. There was no significant correlation between error rates in the hearing screening and error rates in the main experiment (Pearson's r = .078, p = .73).

considered to be the safest (inasmuch the most conservative) strategy, even though one subject explicitly commented after completing the experiment that he mostly perceived /ʃp/ and hardly ever /sp/. All other transcriptions of /sp/-stimuli were retained. Five cases in which the transcription of the onset was not unambiguously interpretable, were also excluded from the analysis.[12] After this procedure, 3400 of the original 3520 observations were left in the data set (exclusion rate: 3.41%).

Logistic regression

A mixed-effects logistic regression was fitted with the lme4 package (Bates et al., 2015) in R (R Core Team, 2016), starting with only fixed effects of German cluster frequency and sonority violation, random slopes for both by subject and a random intercept for stimulus nested under target cluster. In addition to the variables that were relevant for the L1 data, English consonant cluster frequencies were added to the model because of the hypothesis that English-L1 subjects would be influenced by them in addition to the German cluster frequencies. The best fitting model included German log cluster frequency, English log cluster frequency, and their interaction, as well as summed frequency of neighbour clusters and intensity of the onset as numerical fixed effects, and sonority violation as a categorical fixed effect.

In order to specifically test for listener group effects, the L1 and L2 data were analysed in a single logistic regression model with *language group* (L1 vs. L2) as a grouping factor. This model included a three-way interaction between German cluster frequencies, English cluster frequencies, and listener group.

In order to ensure that the inclusion of data from the three subjects who had acquired German before English did not distort the results, the final regression model was run again without the data from these three subjects. The results remained the same.

[12] The full transcriptions were <s-iel>, <7>, <sipch>, <s->, and <s-ein>.

Analysis of misperceptions and confusion matrices

As with the L1 data, misperceptions that constituted legal German CC or CCC clusters were compared to target clusters in terms of cluster frequency and sonority distance.

Confusion matrices for the onset as a whole and for C1 and C2 were set up in the same manner as for the L1 data.

6.6. Results

The overall error rate in the experiment was 38.6%. Error rates ranged from 27% to 59% across participants. For the onset clusters, they ranged from 9% (for /ʃt/) to 94% (for /ks/). Figure 6.2 visualises the error rates over consonant cluster, in addition to the error rates in the L1 group for comparison. The stimulus item /skoːt/ had an intercept > 2.0 on the logit scale for the L2 group, as it did in the L1 group. It was nonetheless included in the data set (see Section 5.5 for reasons). The intercepts of the other stimulus items that were disproportionally high among the L1 group were < 2.0 for the L2 group.

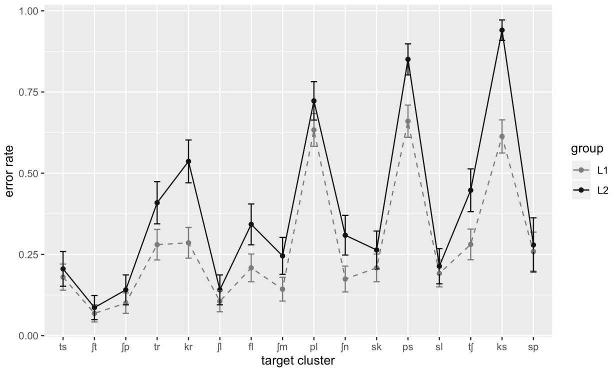


Figure 6.2.: Error probabilities of individual target clusters by group

6.6.1. Logistic regression of L2 data

German log cluster frequency, SSP violation, German neighbourhood frequency, and onset intensity yielded significant main effects. The interaction between German and English log cluster frequencies was also significant. There was no significant main effect of English log cluster frequencies. Table 6.2 displays the estimates, standard errors and z values for the individual predictors.

predictor	estimate	SE	z value	
(Intercept)	-0.32955	0.18496	-1.782	.
German log cluster frequency (type)	-0.85910	0.19344	-4.441	***
English log cluster frequency (type)	0.13010	0.13767	0.945	
SSP violation (ref. level: no violation)	-2.04300	0.41978	-4.867	***
summed neighbourhood frequency (German)	0.26146	0.04976	5.255	***
onset intensity	-0.15593	0.03363	-4.637	***
NAD	-0.39967	0.09271	-4.311	***
German log cluster freq × English log cluster freq	1.14837	0.21692	5.294	***

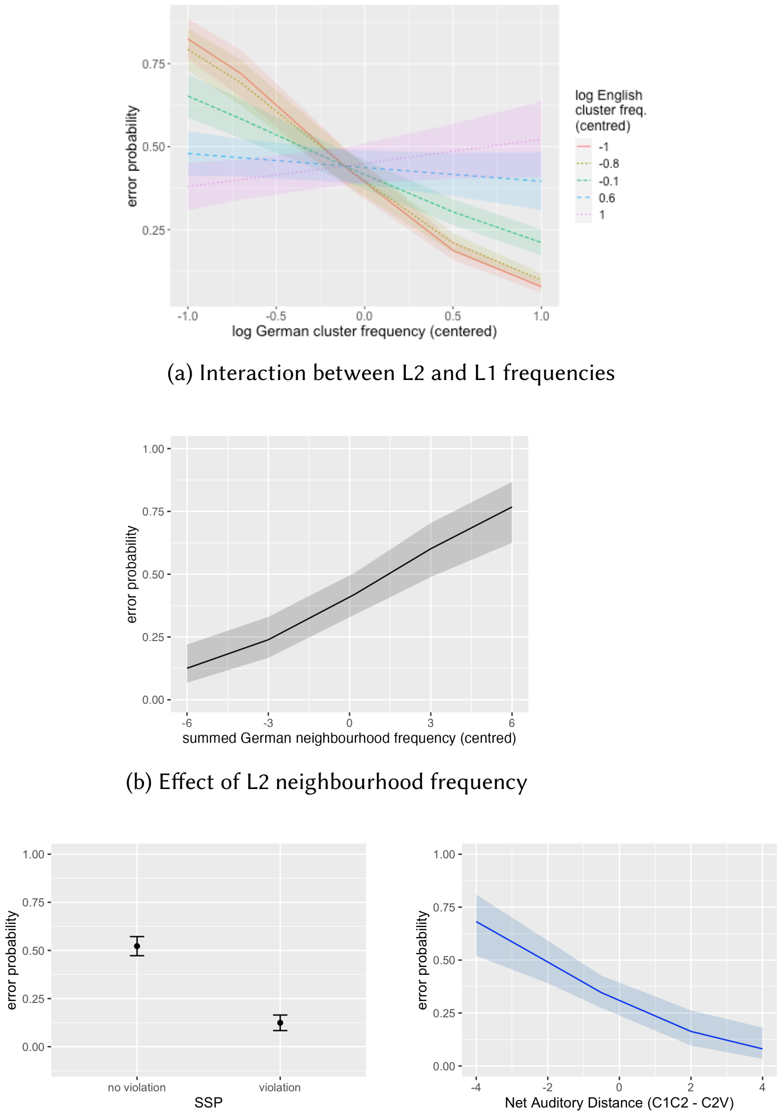
Table 6.2.: Model output of the best-fitting model (data of L2 group only)
Significance codes: *** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1
Formula: `error~logFreqDE*logFreqEN + son.vio + ons.intensity + accNF + NADdiff + (logFreqDE*logFreqEN + son.vio + accNF + NADdiff|subjID) + (1|onset.targ/stimulus)`

The effect of German cluster frequency was as hypothesised: the higher the frequency of a cluster in the L2, the higher the likelihood of correct perception. The interaction with English frequency reveals that this effect was strongest for clusters that have a very low frequency in English (the regression line is steepest for clusters with an English frequency of −1 on the centred scale, see Figure 6.3a). For recognition rates of clusters with a high frequency in English, German frequency was not as relevant. Instead, there was a slight trend in the opposite direction: German-HF clusters had slightly higher error rates than German-LF clusters. At the same time, it can be said that, for clusters with a low frequency in German (the left margin of the German frequency scale), there was a difference in clusters error rate as a function of their L1 frequencies. English-LF clusters had higher error rates than English-HF clusters. However, there did not seem to be an additive effect of L1 and L2 HF clusters, given that they are not the group of clusters with the lowest error rate. The effect of German cluster neighbourhood frequency was also in line with expectations: the higher the summed neighbourhood frequency, the higher the competition for the target, which led to a higher error rate (see Figure 6.3b).

As for the L1 group, SSP violation exhibited an effect that is difficult to account for. Clusters that violate the SSP had a significantly lower error rate than clusters that conform to it (see Figure 6.3c). In contrast to sonority, NAD did show an effect in accordance with phonological theory: the greater the difference in net auditory distance between C1-C2 and C2-V, the better the onset cluster was perceived (see Figure 6.3d).

Intensity of the onset also showed a significant effect, with onsets of higher intensity recognised with greater accuracy.

These findings largely replicated the results of the L1 study: language-specific phonotactics, as well as competition between alternative phoneme sequences in the target language once again proved to be relevant to pseudoword perception, while sonority did not. However, in contrast to the L1 data, the phonological concept of NAD showed a significant effect in L2 perception. The L2 data also indicate that L1 frequencies are not directly relevant to L2 perception (absence of a main effect of English frequencies) but rather modulate the role of L2 frequencies (interaction between English and German frequencies).



(a) Interaction between L2 and L1 frequencies

(b) Effect of L2 neighbourhood frequency

(c) Effect of SSP violation

(d) Effect of Net Auditory Distance

Figure 6.3.: Significant effects in the L2 perception experiment

6.6.2. Logistic regression of combined L1 and L2 data

Table 6.3 shows the output of the regression model with data from both the L1 and the L2 groups. There were significant main effects of German log cluster frequency, summed German neighbourhood frequency, SSP violation, and onset intensity. English log cluster frequency did not show a significant effect. These effects correspond to the ones in the separate models. Furthermore, there were a significant main effect of language group (recognition accuracy was higher for the L1 group) and significant two-way interactions between German log frequency and group and between English log frequency and group.

The interaction between group and sonority violation did not reach significance. Both groups showed higher error rates for clusters that conform to the SSP than for clusters that violate it (see Figure 6.4).

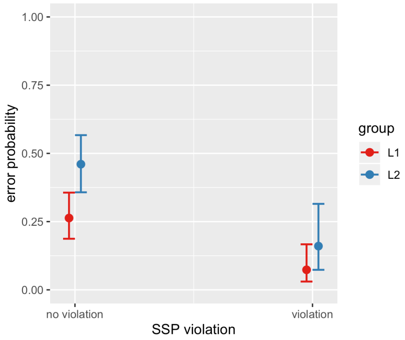


Figure 6.4.: Interaction between sonority violation and language group

There was a significant three-way interaction between German log cluster frequency, English log cluster frequency, and group: for the L2 group, the effect of German cluster frequency increased as a function of lower cluster frequency in English.[13] This was not the case for the L1

[13] This is the same effect as described above for the two-way interaction between German and English cluster frequency in the model for the L2 listener data only;

group. For the Germans, the effect of German cluster frequency was approximately equally strong, irrespective of English cluster frequency (see Figure 6.5).

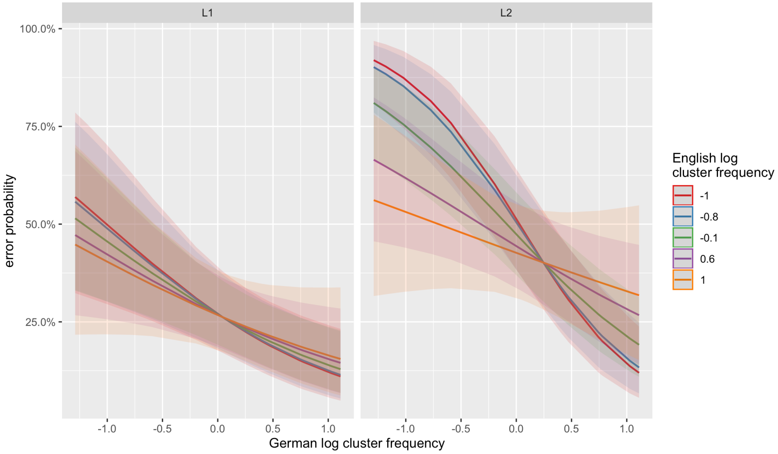


Figure 6.5.: Interaction between German and English cluster frequencies and language group

note, however, that the regression lines are at an angle when compared to the ones in Figure 6.3a due to the fact that the regressors in the two models are not completely identical.

predictor	estimate	SE	z value	
(Intercept)	-0.999571	0.223234	-4.478	***
German log cluster frequency (type)	-0.801654	0.250177	-3.204	**
English log cluster frequency (type)	-0.008158	0.156197	-0.052	
SSP violation (ref. level: no violation)	-1.505583	0.536493	-2.806	**
summed German neighbourhood frequency	0.136397	0.062624	2.178	*
onset intensity	-0.202748	0.037035	-5.475	***
group (ref. level: L1)	0.878987	0.171510	5.125	***
German log cluster freq × English log cluster freq	0.183166	0.259051	0.707	
German log cluster freq × group	-0.330852	0.154398	-2.143	*
English log cluster freq × group	-0.166811	0.073340	-2.275	*
SSP violation × group	0.007855	0.270654	0.029	
German × English log cluster freq × group	0.529098	0.114335	4.628	***

Table 6.3.: Summary of the best-fitting model (data of L1 and L2 group)
Significance codes: *** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1
Formula: `error~logFreqDE*logFreqEN*group + son.vio*group + ons.intensity + accNF + (logFreqDE + son.vio|subjID) + (1|onset.targ/stimulus)`

6.6.3. Analysis of misperceptions and confusion matrices

Analysis of misperceptions

Even among the misperceptions, the vast majority of percepts (97.03%) constituted legal German onsets. A third of the misperceptions were simple onsets.[14] Of all the misperceptions, 61.41% were neighbour clusters of the target according to the definition of deviation by one feature (e.g., /dr/ for /tr/, /ʃm/ for /ʃn/ etc.).

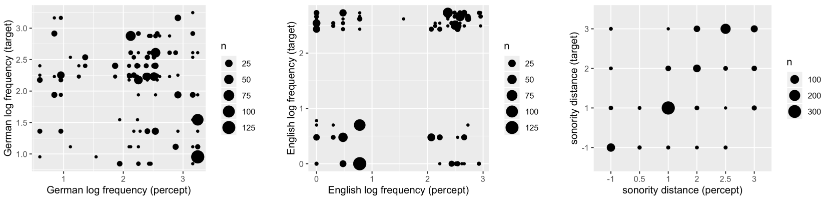
When only examining the cases in which a misperception resulted in another CC onset, a comparison of German cluster frequencies of targets and percepts reveals that 66% of all misperceptions led to a cluster of higher frequency in German. Furthermore, the mean frequency of the reported clusters is higher than the mean frequency of the target clusters (mean log frequency of percepts: 2.60; mean log frequency of targets: 1.97). Hence, in line with the hypothesis, misperceptions im-

[14] For consistency, /ts/ and /tʃ/ are not counted as simple onsets here since they are treated as composed of /t/ + sibilant throughout the whole thesis.

proved the onsets in terms of German frequency in most cases. As with the L1 data, there was a huge difference between HF and LF clusters with respect to the direction of the frequency change: LF clusters were perceptually repaired to higher-frequency clusters more than twice as often as HF clusters. (42% of all HF-misperceptions were “frequency improvements”, whereas 96% of all LF-misperceptions were “frequency improvements”.) Figure 6.6a plots the relationship between German target and percept frequencies graphically. As can be seen, most observations can be found in the right half of the graph, which covers the high-percept-frequency area.

Turning now to English frequencies, a slight tendency for perceptual repairs leading to higher-frequency clusters can be observed. In 55% of all misperceptions resulting in another CC onset, the reported cluster has a higher frequency in English than the target cluster (see also Figure 6.6b). However, the mean English cluster frequency of the target clusters is slightly higher than that of the reported clusters (mean log frequency of reported clusters: 1.47; mean log frequency of targets: 1.51). Also with respect to English frequencies, the picture gets more differentiated when considering HF and LF clusters separately. English-LF clusters were reported as English-higher frequency clusters in 94% of cases, whereas English-HF clusters were reported as English-higher frequency clusters in only 25% of cases. It must nevertheless be kept in mind that the English-LF clusters had extremely low frequencies in English so that there are simply not many lower frequency cluster alternatives that could possibly be reported. The same is not true for the German frequencies, which are distributed more evenly. Therefore, the English frequency patterns are less clear-cut than the German frequency patterns.

Using the same reduced dataset (i.e., only misperceptions that constitute a CC onset), the sonority distance values of target and reported clusters can be compared (see Table 6.4 and Figure 6.6c). According to phonological theory, sonority should rise steeply between C1 and



(a) German log freqs

(b) English log freqs

(c) Sonority distances

Figure 6.6.: Comparison of target and reported cluster characteristics in misperceptions (Dot size represents the number of observations.)

C2; hence the prediction is that sonority distance increased in misperceptions. However, it can be seen that the misperceptions improve the sonority profile of the clusters in only a small minority of cases. In the vast majority of cases, the sonority distance between C1 and C2 remains the same (or is not determinable, for example, because the onset does not have the format CC). The number of cases in which the sonority profile of the cluster deteriorated was almost four times as high as the number of cases in which it improved.

case	no. obs.
$\text{SonDist}_{\text{percept}} > \text{SonDist}_{\text{target}}$	58
$\text{SonDist}_{\text{target}} > \text{SonDist}_{\text{percept}}$	223
$\text{SonDist}_{\text{percept}} = \text{SonDist}_{\text{target}}$	519
$\text{SonDist}_{\text{percept}}$ not determinable	511

Table 6.4.: Sonority improvements vs. deteriorations in misperceptions

Confusion matrices

Table 6.5 shows the confusion matrix for the whole onset, Tables 6.6 and 6.7 the confusion matrices for C1 and C2, respectively.

Table 6.5.: **Confusion matrix for consonant clusters in L2 perception**
Rows show target clusters, columns listeners' responses (order of consonant clusters follows token frequencies); columns *C1* and *C2* report cases in which only one of the component consonants was reported as a simple onset, column *voice* reports voicing errors in at least one of the consonants, and column *other* sums up all remaining confusions; numbers represent percentage; note: value for /sp/ > /ʃp/ confusions missing because these cases were excluded from the analysis)

	ts	**ʃt**	**ʃp**	**tr**	**kr**	**ʃl**	**fl**	**ʃm**	**pl**	**ʃn**	**sk**	**ps**	**sl**	**tʃ**	**ks**	**sp**	**C1**	**C2**	**voice**	**other**
ts	79.0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5	16.4	0	3.7
ʃt	0	91.4	6.8	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	1.4
ʃp	0	1.8	85.9	0	0	0	0	0.5	0	0	0	0	0	0	0	5.0	0	0	0	6.8
tr	0	0	0	59.1	0.5	0	0	0	0	0	0	0	0	0	0	0	2.7	0.5	27.3	10
kr	0.5	0.5	0	1.8	45.0	0	0	0	0	0	0	0	0	0	0	0	10.9	3.6	24.5	13.2
ʃl	0	0.5	0	0	0	85.9	2.7	0	0	1.4	0	0	5.0	0.5	0	0	1.4	0.5	0	2.3
fl	0	0	0	0	0	6.4	64.8	0.9	3.2	0	0	0	1.4	0	0	0	0.5	5.0	0.9	16.9
ʃm	0	0	0.5	0	0	0.9	0	75.5	0	3.6	0	0	0	0	0	0	0	5.5	0	14.1
pl	0	0	0	0	0.5	1.4	12.7	0	27.7	0	0	0	0	0	0	0.5	0.9	11.8	11.8	32.7
ʃn	0	0	0	0	0	0	0.5	18.2	0	69.1	0	0	0	0	0	0	1.8	2.3	0	8.2
sk	0.5	4.1	9.5	0	0	0	0	0.5	0	0	73.2	0	0	0	0	5.0	0.5	0.9	0.9	5.0
ps	42.6	0.5	0	0	0	0	0	0	0	0	0.5	13.9	0	0	0	0	0	32.4	0	10.2
sl	0.5	0	0.5	0	0	9.5	4.1	0	0	0.5	0	0	74.1	0	0	0	0.5	0	0	10.5
tʃ	2.7	0.5	0	5.5	0	0	0	0.5	0	0	0	0	0	55.3	0	0	7.3	22.8	0	5.5
ks	59.6	0.5	0	0	0	0	0	0	0	0	0	0.5	0	0	5.0	0	0	27.1	2.8	4.6
sp	0	0.9	–	0	0	0	0	0	0	0	7.2	0	0	0	0	70.3	0.9	1.8	0.9	18.0
sum	185.4	101.2	103.2	66.4	46.0	104.1	84.8	96.1	30.9	74.6	80.9	14.4	80.5	55.8	5.0	81.3	27.9	130.6	69.1	163.1

Table 6.6.: L2-Confusion matrix for C1

	–	p	t	k	b	d	g	h	f	s	ʃ	v
p	22.6	22.8	22.4	10.4	7.1	0.7	0.7	0.2	9.4	0.7	0.9	1.2
t	13.4	0.5	72.3	0.3	0.5	9.6	0.6	0.6	0.9	0.3	0.5	0.3
k	15.5	1.1	32.0	31.7	0.7	1.4	13.9	0	1.8	0.9	0.5	0.2
f	5.0	4.6	0	4.6	2.3	0	0.9	0	71.2	2.3	7.8	0.9
s	0.7	0.2	0.2	0	0.9	0	0.2	0	2.4	83.5	12.0	0
ʃ	1.6	0	0.1	0.6	0.5	0	0	0.2	0.8	5.8	90.0	0

Table 6.7.: L2-Confusion matrix for C2

	–	p	t	k	b	d	g	f	s	ʃ	v	m	n	l	r
p	5.7	86.7	1.5	2.7	0.3	0	0	0	0	0	1.5	0.3	0	0	1.2
t	0	7.3	92.3	0	0.5	0	0	0	0	0	0	0	0	0	0
k	4.1	14.5	5.0	74.5	0	0	0.9	0	0	0	0	0.5	0	0	0.5
s	3.4	0	0.3	0.2	0	0	0	0.2	95.9	0	0	0	0.2	0	0
ʃ	10.0	0	0	0	0	0	0	0	3.2	78.5	0	0.5	0	0	6.4
m	0.5	0.5	0	0	0	0	0	0	0	0	0.5	92.3	5.5	0.9	0
n	3.2	0	0	0	0	0	0	0	0	0	0.5	20.0	75.9	0.5	0
l	4.6	0.1	0.1	0	0	0	0	0	0.1	0	0.1	0.8	1.6	92.2	0.5
r	12.0	0	0.2	0	0	0	0	0	0.5	0	0	0	0.2	0	86.8

As can be seen in Table 6.5, recognition accuracy varied considerably between onset clusters, with a recognition rate above 90% for the easiest cluster, /ʃt/, and a mere 5% for the hardest cluster, /ks/. Since the error rate of 95% for /ks/ was unexpectedly high, it was checked whether all ten stimuli contribute equally much to it. Half of them (/ksɛp/, /ksɪm/, /kso:f/, /ksɔp/, and /ksɔʏl/) were misperceived by all subjects. The others were misperceived by 17 to 21 out of 22 subjects. The five stimuli with a 100% error rate were reinspected auditorily in quiet. No abnormalities could be detected. The spectrograms of the onsets (i.e., the /ks/ parts) of these five stimuli can be found in Appendix A. Like the L1 subjects, the L2 subjects, too, reported target /ks/ more often as /ts/ than correctly. The reverse, whereby /ts/ would be reported as one of its neighbouring clusters, did not occur at all. This was the strongest asymmetry in the confusion matrix. Another interesting asymmetry concerns false alarms: while the HF onsets /ʃt/ and

/t͡s/ attracted responses from nine and six other target clusters, respectively, the LF cluster /ks/ did not elicit any false positives, and /ps/, /pl/, and /tʃ/ each attracted responses from only one other cluster. Hence, low recognition rates were generally mirrored by low false alarm rates.

It is also noticeable that confusion between two consonants depended on the contexts in which the two consonants appeared: in the sibilant + /l/ context, the confusion /s/ > /ʃ/ was almost twice as frequent as the confusion /ʃ/ > /s/, while in the /t/ + sibilant context, the former did not occur at all, but the latter did. However, not all phonemes were equally influenced by context. While there was a relatively small difference in recognition rates for /t/ in C1 or C2 position (74% vs. 92%), the discrepancy was massive for /p/ (25% in C1 position vs. 92% in C2 position).[15]

6.7. Discussion

6.7.1. Acoustic factors

To a large degree, the results for the L2 listeners resembled those of the L1 listeners. In both groups, acoustic factors evidently play an important role in speech perception in noise. Onset intensity had a significant effect on identification performance in the L2 group, and, with the exception of /ts/, the stop-initial clusters had the highest error rates—they ranged from 41% for /tr/ to as much as 95% for /ks/. The reasons for the perceptual disadvantage in perceiving stop–consonant sequences were discussed in detail in Section 4.3. In short, the acoustic cues for stops are masked to a large degree by an ensuing sibilant. Obviously, this reduced perceptibility of stops posed even bigger difficulties for the L2 listeners than for the L1 listeners, as seen in the overall

[15] Since the cases in which /sp/ was transcribed as <sp> were excluded from the analysis, the number for /p/ in C2 position is not completely reliable, but it seems fairly plausible since it lies in the same range as the rates for the similar clusters /ʃp/ and /sk/.

higher stop–consonant error rates for the L2 listeners. Extremely high error rates in the perception of /ps/ by English-L1 listeners have already been reported by Albright (2007b). In this case, it could be argued that learner listeners have problems perceiving a sequence that is illegal in their L1. However, in the present data, the error rates for /pl/, /tr/, and /kr/ were also disproportionately high, which makes an acoustic explanation or at least a partial influence of acoustics more plausible. Furthermore, it is obvious that for the L2 listeners, there was a perceptual advantage for sC clusters, too. Throughout the whole frequency range, they were recognised with an above average probability. As was discussed in Section 4.3, this can be best explained by sibilants' acoustic properties, especially their noise resistance.

A comparison of the confusion matrices in Section 6.6.3 with those from studies that minimise the influence of phonotactics (because they use CVC or VCV nonce syllables) reveals a lot of parallels. In much the same way as acoustically-oriented studies, voiceless stop consonants—with the exception of /t/—had the highest error rates (Lecumberri & Cooke, 2006; e.g.,) and were frequently confused with other voiceless stop consonants (e.g., Lecumberri & Cooke, 2006; Marchegiani & Fafoutis, 2015; for L2 perception). Likewise, when averaged over all consonant classes, place of articulation confusions constituted the most frequent misperceptions both in the present study and a number of acoustically-oriented studies (Benkí, 2003; Lecumberri & Cooke, 2006; Moreno-Torres et al., 2017; Woods et al., 2010). This suggests that, on a purely acoustic basis, place of articulation information is less resistant to noise than manner or voicing information (cf. Lecumberri & Cooke, 2006; for a similar interpretation).

For some of the observations, it is difficult to interpret whether they are acoustically grounded or whether they have their origin in listeners' knowledge of the language structure. For example, voiceless stops are relatively frequently confused with their voiced counterparts in C1 position, but voicing errors hardly occur on stops in C2 position. This could be due to the more favourable prevocalic environment in which the silence before voice onset is clearly recognisable and the release burst, whose intensity helps differentiate between voiceless and voiced stops, is unimpeded. However, it also correlates with the legality of the potential clusters that misperceptions would produce. With a stop in C1 position, a voicing error would produce the legal clusters /dr/, /gr/, and /bl/ (and Table 6.5 confirms that this is exactly where the voicing errors occur). Conversely, a voicing error would not turn any of the clusters with a stop in C2 position (/ʃt/, /ʃp/, /sk/, /sp/) into a legal German onset cluster. Thus the larger number of voicing errors on stops in C1 position could also be informed by the listeners' knowledge about the legality of the outcome.

6.7.2. L2 phonotactics

The analyses showed that L2 perception was strongly influenced by gradient L2 phonotactics: there was a significant effect of German cluster frequency in the L2 group. This supports the hypothesis that target-language phonotactics influences speech perception in L2 listeners as well. In fact, the German frequency effect was even stronger for the L2 listeners than for the native listeners, as can be seen in the two-way interaction between German frequency and language group. The L2 phonotactic effect is in line with earlier studies that demonstrated that L2 listeners are able to make use of the structural characteristics of the target language during L2 processing (e.g., Carlson et al., 2016; Hanulíková et al., 2011; Lentz, 2011; Trapman & Kager, 2009; Weber & Cutler, 2006). In contrast to many previous studies, the results of the study at hand show that it is not only legality knowledge about the L2 that is acquired by the learners but gradient phonotactic knowledge in the form of consonant cluster frequencies. This means that statistical learning is not restricted to the L1 but also operates in languages acquired later and is probably language-specific; this suggests that phonotactic distributions are acquired separately for each language and are labelled accordingly (cf. Lentz, 2011; p. 164).

The more extreme effect in L2 listening may seem surprising at first but can be explained by a skewed distribution of phonotactic sequences in the mental lexicon: the relative frequency difference between the most and least frequent consonant clusters is probably greater for the L2 listeners than for the native listener group. Even with their limited German input, they will have heard the most frequent German phoneme sequences many times, but they may not have heard the least frequent ones at all, which makes infrequent consonant clusters illegal according to their internal representations. For L2 learners, therefore, the frequency distributions themselves are probably more extreme, which leads to a more extreme frequency effect. This is in line with the conclusion by Hanulíková et al. (2011; p. 516) that learner frequencies are determined by "[...] a subset of L2 [structures] (most likely the more frequent structures [...])". For exemplification, consider the case of initial stop + /s/ sequences: it is striking that the L2 listeners showed close to native-like performance in the perception of /ts/, while their error rates for /ps/ and especially /ks/ were higher than the already high error rates for those clusters in the L1 group. This strong discrepancy between the stop–fricative onsets probably stems from the listeners' diverging experience with such sequences: while they have encountered initial /ts/ in many German words (e.g., *zwei* "two", *zehn* "ten", *zu/zurück* "to/back", *Zeit* "time", *Zahn* "tooth", *Zug* "train", *Zahl* "number" etc. from the core vocabulary), they might never have heard the sequences /ps/ and /ks/. The same can be said about the frequency difference between other onset clusters. While there are more than 200 words starting with /ʃt/ that have a frequency above 20 in the Mannheim Corpus, and more than 100 starting with /ʃp/, there are none starting with /tʃ/, /sp/, or /sl/ (values are taken from the WebCELEX German Lemma Database, http://celex.mpi.nl/). Consequently, since the L2 listeners have reduced input in German compared to the L1 group, the relative frequency difference between the HF clusters and the LF clusters is larger for them. This calls into question the applicability of the frequency measure used here (German cluster frequencies based on L1 lexicon entries) for L2 learners. If they have encountered a skewed phonotactic distribution, a more accurate basis for their mental representations of German sublexical frequencies would probably be learner corpora (for token frequencies) and lexicons (for type frequencies), paralleling the situation in L1 acquisition research in which frequency counts for psycholinguistic experiments are also usually derived from child language corpora. However, given this drawback, the significant effect of German cluster frequencies is all the more remarkable. It implies that using an L1-based count for German frequencies is indicative of how far the intermediate to advanced learners have already come on the road to native-like frequency representations.

What probably also contributes to the discrepancy between HF and LF clusters is L2 listeners' lack of phonetic knowledge on how to parse the cues for clusters they have never or only rarely encountered (cf. Davidson & Shaw, 2012). Without sufficient exposure to appropriate stimuli, listeners cannot acquire knowledge about the relevant phonetic cues for phoneme identification in a specific context, for example, identification of initial stops before fricatives.

In addition, L2 listeners might also rely on top-down information to a larger degree than L1 listeners during listening due to their inferior ability to interpret the acoustic signal. Based on a stronger lexical neighbourhood effect for L2 than for L1 listeners (cf. Bradlow & Pisoni, 1999; Marian et al., 2008), Cutler (2012; pp. 360–361) remarks: "L2 listeners are highly susceptible (to an even greater extent than L1 listeners) to the precise makeup of their vocabulary." This can, by extension, also be said for the sublexical level: listeners use top-down information available to them in order to interpret the acoustic signal, and this includes previously acquired knowledge about language-specific phonotactic distributions.

6.7.3. L1 phonotactics

However, the influence of L2 phonotactics cannot be contemplated without reference to L1 phonotactics. The interaction between German and English cluster frequencies (and, likewise, the three-way interaction between German cluster frequencies, English cluster frequencies, and group in the models for both L1 and L2 listeners) reveals that the strong inhibitory L2 phonotactic effect emerges only in clusters that occur very rarely or not at all in the L1. A possible interpretation of this finding is that experience with the clusters from the L1 "evens out" the experience in the L2 or, to put it the other way around, learning the phonotactic distributions of a new language is easier when the learner can start with a *tabula rasa*. What is surprising in that respect is the lack of a main effect of English frequencies. If there is an effect of German frequencies on English-illegal clusters but for English-legal clusters there is not, then it would be expected that their recognition is influenced by their English frequencies instead. However, that is not what can be seen in the data. If anything, a slight trend for L1-frequent clusters to lead to higher error probabilities can be seen (cf. Figure 6.7).

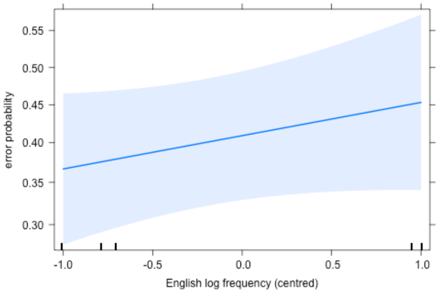


Figure 6.7.: L1 frequency effect (n.s.)

In the model that compares the two listener groups, on the other hand, there was a significant interaction between listener group and

English cluster frequencies, which shows a decrease in error rates as English frequencies increase for the Australians but not for the Germans. This difference is hard to account for solely in terms of the additional regressor present in the model with only L2 data (NAD difference). Since the effect of English frequencies is part of a higher-order term in both models, it should be interpreted with caution, however.

In order to investigate whether a main effect of English frequencies in the existing clusters was obscured by the presence of marginal to illegal clusters, the dataset was reduced to the native English clusters, and the model was run again without the added interaction of German clusters. The model output (see Table A.5 in the Appendix) did not show the expected effect of English frequencies. There was, however, a trend in the expected direction. Whether or not the distribution of frequencies in English was even enough for a clear effect to emerge is worthy of further consideration. As mentioned, there were no clusters from the English middle-frequency range (see also Figures 6.1 and 6.7). In this case, a categorical measure of L1 phonotactics (i.e., English cluster legality) might capture the situation better. To test this possibility, a model with English legality instead of English frequency was fitted (all other predictors remained the same as in Table 6.2). Just like the model with gradient L1 phonotactics, it showed no main effect of L1 phonotactics but a significant interaction between L1 and L2 phonotactics (model summary in Table A.6 in the Appendix). A comparison of both models—the original one, which features English cluster frequencies (numeric/continuous), and the new one, which features English cluster legality (binary) instead—showed very similar values of prediction error; the difference in both the Akaike information criterion and the Bayesian information criterion is < 1, with slightly lower values assigned to the numeric model (AIC: 3295.3 vs. 3295.8; BIC: 3528.3 vs. 3528.8). If any model preference can be based on these numbers, it would speak in favour of the numeric model. More importantly, however, the interaction with German cluster frequencies showed a gradual effect of English frequencies: the slopes for the effect of German frequencies were different for all values of English frequencies, and they form an ordered pattern (see Figure 6.3a on p. 178). This suggests that the effect of L2 frequencies is gradual rather than categorical; nevertheless, it is not a main effect but an interaction.

These findings are in line with a number of studies that report effects of L2 (i.e., target-language) phonotactics but not of L1 phonotactics (Boll-Avetisyan, 2011; Lentz & Kager, 2015; Lentz, 2011; ch. 4). Although these studies did not test for an interaction of L1 and L2 frequencies, Boll-Avetisyan (2011; p. 10) notes that learner performance was "native-like [i.e., affected by L2 biphone frequency] and not affected by whether the syllable types were attested in the L1", which means that she tested for an interaction between *gradient* L2 and *categorical* (structural) L1 phonotactics, without finding an effect.

How do the results from the present study relate to previous findings on perceptual illusions in L1-illegal structures? In this experiment, there was no main effect of L1 phonotactics, which means that not *all* L1-illegal structures were repaired. However, Trapman and Kager (2009) note that the repair of L1-illegal structures depends on their degree of markedness. Interestingly, this is the pattern found in the L2 data if markedness is defined not in a universal sense (which is probably what the authors meant) but as marked structure within the L2. Of the consonant clusters that are illegal in English (the L1), only those are subject to perceptual illusions that are also marked (i.e., marginal) in German, the L2: /ks/ and /ps/. Their occurrence is mainly restricted to Greek loan words. All the other L1-illegal clusters (/ts/[16], /ʃt/, /ʃp/, /ʃl/, /ʃm/ and /ʃn/) seem to be unmarked enough in German to be faithfully perceived in most instances. In other words, their frequency in German

[16] In the case of initial /ts/, support may also come from words of foreign origin that are used commonly enough in English, such as *tsunami*, *tsetse (fly)*, *tsatsiki*, or *Tswana*. Thus, although this cluster does not occur in syllable-initial position in native English vocabulary, it might not be perceived as illegal by English speakers.

is high enough to facilitate their perception, while that of /ks/ and /ps/ is too low, which leads to a substantial number of perceptual illusions. Therefore, instead of a repair of all L1-illegal sequences, a repair of only those that are L2-marked structures in the L2 can be observed. These are probably considered illegal by many of the L2 learners, who have not encountered them in either the L1 or the L2.

An alternative interpretation refers to the status of these clusters in the L1 alone. Note that studies reporting perceptual illusions tested structures that are completely illegal in the listeners' L1—and which the listeners did not know from any other language in most cases. In the present study, only two of the clusters, /ps/ and /ks/, are structurally illegal in the L1 since English does not allow initial stop–fricative clusters (although cf. Section 6.3 for marginal use). As mentioned in Section 6.3, all other clusters that do not exist in English, that is, clusters with /ʃ/ in C1 position, have English equivalents with /s/ in C1 position. Therefore, the structure sibilant–stop/nasal/liquid is not illegal in the L1. Indeed, Albright (2009; p. 23) also refers to initial clusters, such as /ʃl/, as "English-like". Likewise, the well-recognised /t͡s/ has the English equivalent /t͡ʃ/. Consequently, it is only clusters that are *structurally* illegal in the L1 that reliably caused perceptual illusions in L2 listeners. They were confused with a single competing cluster more often than they were correctly identified.[17] However, it is unclear which of the two factors—marginality in the L2 or complete structural illegality in the L1—is the decisive factor for perceptual repair to occur.

Carlson et al.'s (2018, 2016) findings that perceptual repairs of L1-illegal sequences decline with growing proficiency in a language that does allow them are relevant in answering this question. If the repairs of /ps/ and /ks/ in the present data are caused by their structural illegal-

[17] Some of the L1-legal clusters also had high error rates, but these were still far below those of L1-illegal structures and cannot be attributed to a single competing cluster attracting more responses than the target.

ity in English, then beginner and advanced learners of German should show an equal tendency to repair these clusters. If, on the other hand, this is primarily due to their marginality in German, then it can be hypothesised that advanced learners have already acquired their legality, while less advanced learners have not. To resolve this matter, the L2 listeners were divided into an intermediate (CEFR levels B1 and B2) and an advanced (CEFR levels C1 and C2) learner group, and an interaction between learner group and structural legality was tested in a logistic regression model. As Table 6.8 shows, there was no significant interaction between L2 proficiency and L1 structural illegality. The error rates for structurally illegal clusters were even slightly, though not significantly, higher in the advanced group (see Figure 6.8). This goes against the hypothesis that advanced learners will have acquired the legality of these sequences and suggests that it is the structure's illegality in English that leads to the extremely high error rates. Even though advanced learners have had a high amount of German input, they have not acquired the legality of these clusters and thus are prone to much more consistent perceptual illusions with regard to these clusters than the other LF clusters. The discrepancy between the results of Carlson's studies in which increasing L2 proficiency helped learners to acquire L1-illegal structures and overcome perceptual illusions, and those of the present study is due to the diverging frequency or markedness status of the structure in L2: while Carlson (2018) and Carlson et al. (2016) tested structures that are common in the L2, the structure that led to perceptual repairs in the present study is only very marginal in the L2.

To conclude this section on language-specific phonotactics, a short note regarding the gradient vs. categorical nature of the frequency effect should be given since so many previous studies treated phonotactics as a categorical variable. It is clear that the L2 phonotactic effect is gradient since only legal German consonant clusters were used that differ in their frequency of use, and this difference had an influence on the listeners' perception. The significant interaction between English

predictor	estimate	SE	z value	
(Intercept)	-1.086495	0.298687	-3.638	***
L2 proficiency (reference level: intermediate)	-0.008998	0.249405	-0.036	
struct. illegality (reference level: legal)	3.316766	0.730806	4.539	***
L2 proficiency × structural illegality	1.294335	0.388440	3.332	***

Table 6.8.: Model testing interaction between learner group and L1 structural illegality
Significance codes: *** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1
Formula: `error~L2group*str.illegality + (L2group*str.illegality|subjID) + (1|onset.targ/stimulus)`

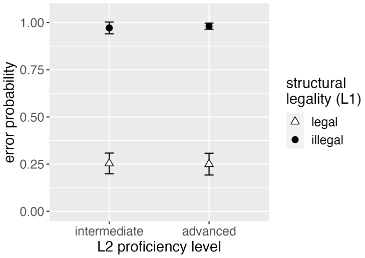


Figure 6.8.: Interaction between L2 proficiency and cluster structure legality in English

and German cluster frequencies shows that L1 influence is gradient as well. The model that uses English legality as a predictor rather than English frequency showed the same effect and did not capture the data better. It most likely produced an aggregation of gradient changes into two categories and consequently a simplification of reality. It therefore seems that the finer measure, frequency, is the more psychologically realistic one. In studies that test for only categorical phonotactic influences, however, the resulting effect might appear to be one of legality. It would be interesting to repeat the present experiment with clusters that cover a wider range of English frequencies. The results should not differ from those at present.

6.7.4. Sonority and Net Auditory Distance

As in the L1 data, SSP violation of an onset cluster did not yield the expected effect of reduced perceptibility but, on the contrary, led to a perceptual advantage. Clusters that violate the SSP were correctly recognised more often, and for the majority of misperceptions, the sonority distance between the two consonants of a cluster remained unchanged. When it is changed, it causes a reduction in sonority distance between C1 and C2 almost four times as often (relative to C2 and V) as an increase. Since no phonological theory of sonority would predict this, it is clear that the facilitating effect of SSP-violating clusters must have its origin in a different principle that is correlated with sonority. As discussed above, it is probably the perceptual advantage of sC clusters. The L2 listeners thus were no more affected by sonority sequencing than the L1 listeners.

The results of this study contrast with those from Ulbrich and Wiese (2018), whereby conformity of consonant clusters to the SSP is more important in L2 word learning than their L2-legality. The two studies differ not only in terms of the task (identification in noise vs. recollection of word–picture pairs) but also in the composition of test clusters: while all SSP-violating clusters in the present study are legal in German, Ulbrich and Wiese (2018) used a crossed design of L2-legality and SSP-conformity. Therefore, the subjects in the present study may have been more inclined to rely on language-specific phonotactics, which does not serve as reliable guidance in Ulbrich and Wiese's study. Alternatively, the diverging results could be an indication that sonority takes on a more important role in recollection and learning than in perception. This is plausible considering that perception is mainly determined by acoustic factors, which—as discussed above—run counter to the effects of sonority in cases like the present. Support for the view that sonority sequencing is more important for more conscious tasks than perceptual processing comes from Trapman and Kager's (2009) study in which a sonority effect was found in word-likeness ratings but not in lexical decision. Compared to the present study, they used a broader range of SSP-violating clusters (stop–stop, liquid–stop, fricative–stop), only one of which contained a sibilant in C1 position. Hence their (null) result in the processing task is less influenced by the good perceptibility of sibilant–stop clusters than the "anti-sonority effect" found in the present study and thus probably more realistic with respect to the true influence of sonority sequencing on L2 perception.

Based on the above considerations, it can be assumed that L2 listeners are no more influenced by sonority sequencing in perception in noise than L1 listeners. Nevertheless, the results for Net Auditory Distance suggest that this finding cannot be generalised to all universal phonological principles. NAD yielded a significant effect such that clusters with a greater NAD difference between C1 and C2 (relative to that of C2 and V) had lower error rates. Thus, the influence of NAD on perception is as predicted by phonological theory (specifically, Beats-and-Binding Phonology; cf. e.g., Baroni, 2014; Dziubalska-Kołaczyk, 2007). The superiority of NAD over sonority in accounting for the results of the present study are not surprising given that the concept of NAD (and B&B phonology in general) has a strong foundation in articulatory and perceptual observations. The results of the present experiment support the view that phonological principles not related to psycholinguistic processes have little value in accounting for psycholinguistic phenomena. Their application is probably restricted to the core area of abstract language description. It is not surprising that, when investigating speech perception, turning to phonological principles that are more extensively informed by speech perception, like NAD, is more fruitful. Strangely, no study (to the best of my knowledge) has yet examined the influence of NAD on the perception of consonant clusters. The present investigation shows that this is a promising direction for further research.

6.7.5. Comparison with L1 group

Unsurprisingly, the overall error rate was higher for the Australian listeners than for the native listeners (41% vs. 31%). Nonetheless, a glance at the error rates for the individual clusters (cf. Figure 6.2 on p. 175) reveals that the two groups behave extremely similarly in terms of which clusters are difficult to identify and which are not.

The results of the regression models (one model comprising data from both listener groups, as well as a comparison of the individual models) confirm that there are only two predictors that differ in their effect between the two groups: 1) The effect of German cluster frequencies is modulated by their English frequencies in the Australian group, which is naturally not the case for the German group, and 2) the L2 listeners show an effect of NAD. The strong correspondence in the data between the two listener groups shows that, in principle, L2 listeners are susceptible to the same influences as L1 listeners. First of all, they are sensitive to cluster frequencies in the target language. This suggests that they are able to employ distributional knowledge about the *target* language and are not misled by the frequencies of their L1, which are irrelevant in L2 listening situations. The target language frequency effect is even stronger for the L2 group than for the L1 group. This parallels the results of a reading study by Lemhöfer et al. (2011), who found a stronger orthotactic effect for L2 readers than for L1 readers. Similar results have also been obtained for production (repetition accuracy) when comparing children during L1 acquisition to adults (Edwards et al., 2004). In this case, the effect of sublexical transitional probabilities was more extreme for the L1 learners than for the adults. However, the interaction with L1 frequencies indicates that L1 phonotactics still has an influence on L2 perception, albeit an indirect one. Second, there was an inhibitory effect of cluster neighbourhood frequencies for both listener groups, which means that the same interplay between activation and competition of target-language units is in force for L1 and L2 listeners.[18] Moreover, none of the listener groups displayed the expected sonority effect, but, instead, a significant effect in the opposite direction was found. In summary, the two listener groups behave remarkably similarly in terms of what their perception of consonant clusters is influenced by.

However, there are a number of interesting differences between the two groups. The most obvious one is the NAD effect for the L2 group, which did not emerge in the L1 group. One reason could be that the acoustic cues are more difficult to identify and interpret for the non-native listeners. As a result, they are likely to benefit more from a greater auditory distance between two consecutive phonemes, which results in greater contrastivity. The native listeners, on the other hand, probably do not depend as much on this kind of auditory modulation because they are more experienced in interpreting acoustic cues, even for less contrastive phoneme sequences. In effect, this could be an indication that universal principles, when based on the relevant parameters, are indeed more influential in L2 perception than in L1 perception. This calls for more research on the potentially dissimilar role of universal principles in L1 perception and L2 perception.

Secondly, the Australian listeners made considerably more voicing errors in all three clusters which have proven to be prone to this kind of misperception (/tr/, /kr/, and /pl/; see confusion matrices on pp. 115 and 184 for comparison). As Kwon and Chitoran (2019; p. 388) note, the "perception of foreign consonant clusters is guided not only by native phonotactics, but also by preferred inter-consonant timing patterns of the listeners' native language". A plausible explanation is that it is differences in timing patterns of German and English that lead to the mis-

[18] Numerous studies (e.g., Blumenfeld & Marian, 2013; Marian & Spivey, 1999; Spivey & Marian, 1999; Weber & Cutler, 2004) have shown that, in bilinguals, lexemes from both languages are activated in parallel during perception, which leads to greater competition. It would be interesting to test whether in the present case additional English neighbouring clusters also add to the competition.

perception of voiceless consonants as voiced. In German stop–liquid sequences, contrasts between voiced and voiceless stops are strongly reduced because, even for voiced stops, voice onset occurs only shortly before or in parallel with closure release (Kohler, 1977). Moreover, the aspiration that usually distinguishes voiceless from voiced stops is missing before liquids in German (cf. Kohler, 1977) but not in English (cf. Scherer & Wollmann, 1986; p. 86). This shows how different phonetic realisations can be a source of errors for L2 listeners (cf. also Davidson & Shaw, 2012). The high number of voicing errors for clusters that have a legal voiced counterpart but *not* for clusters in which an illegal cluster would result is another impressive demonstration of how L2 listeners deploy their phonotactic knowledge of the L2 to compensate for their weaker interpretive abilities of acoustic cues.

Lastly, the tendency for perceptual repair of a marginal cluster to a common one was more pronounced for the L2 listeners. The misperceptions of /ks/ and /ps/ as /ts/ have already been discussed in depth. The same can be seen in a number of other cluster confusions (see Table 6.9 for a short overview). Take, for example, the case of /sl/ > /ʃl/—this misperception, too, is far more common in the L2 group than in the L1 group. The absence of /s/ > /ʃ/ confusions in the context of target /ts/ indicates that this is not due to misinterpretation of acoustic cues. Rather, what can be seen here probably demonstrates the impact of expectation based on phonotactic knowledge. In the case of L2 listeners, it probably also involves some sort of hypercorrection: they most likely know that /sl/ is not part of the native German repertoire and that English /sl/ corresponds to German /ʃl/ (e.g., *sleep/Schlaf*, *sling/Schlinge*, *slumber/schlummern*). The native listener group was less reluctant to report hearing /sl/. This may, on the one hand, be due to their better interpretation of the acoustic cues, but the interpretation may also have received more lexical/phonotactic support for the L1 listeners since they know that /sl/ can occur in German speech in words of foreign origin, like *Slang*, *Slalom*, or *slawisch*. For the L2 group, this cluster is probably labelled as belonging to English phonotactics because they are very familiar with it from English words but may not have come across this sequence in German words.

repair	L1 group	L2 group
/ks/ > /ts/	43.8	59.6
/ps/ > /ts/	28.9	42.6
/sk/ > /ʃp/	6.6	9.5
/sl/ > /ʃl/	0.6	9.5
/tʃ/ > /ts/	1.1	2.7

Table 6.9.: Perceptual repairs from marginal target clusters to percepts of common clusters
(Numbers denote percentage of all responses for a target cluster.)

The tendency for hypercorrection on the part of learners can also be seen in the relatively high number of /tr/ responses to target /tʃ/, which do not occur in the L1 group at all. In English, the realisation of the cluster /tr/ is acoustically very close to [tʃ][19] (e.g., realisation of /traɪ/ as [t͡ʃʰɹaɪ]) because the release of the stop involves friction caused by the curling back of the tongue in assimilation to the following [ɹ] (Ogden, 2009). The Australian listeners seem to have tried to compensate for this by ascribing the realisation [tʃ] to an "underlying" /tr/, even though this assimilation process does not take place in German. Their interpretation was probably also influenced by their knowledge that /tr/ is a common German onset, while /tʃ/ is not. Even though the latter occurs in a few common words, like *tschüs* and *ciao*, both its token frequency and—more importantly—its type frequency are low, certainly much lower than those of /tr/.

Lastly, hypercorrection is demonstrated by the percentage of illegal percepts: less than 3% of L2 listeners responses contain illegal syllable onsets, whereas for the native listeners, the number is more than twice as high.

[19] According to Ogden (2009; p. 110), it can even be classified as an affricate.

6.7.6. Effects of language mode

In the experiment, the Australian subjects were deliberately put into German language mode. The introduction and instructions were given in German, and participants were told that they would hear nonce words that sound German. They were also instructed to write down their responses according to German spelling conventions. It is well-known that expectations are important in speech perception (e.g., Brown & Hildum, 1956; Hawkins, 2010), so the fact that they expected German syllables most probably influenced language activation and hence the processing of the stimuli. The fact that the data show a clear effect of target-language frequencies but no direct influence of native frequencies supports theories about language mode according to which the target language is more activated during speech processing and is given priority in cases of conflicting perceptual hypotheses. It also replicates previous findings (e.g., Freeman et al., 2016; Lentz, 2011) that, in addition to the general frame of the experiment (language of introduction and instructions), which causes expectations about the input, the composition of the stimuli plays an important role for the activation of the target- vs. the non-target language. It has been found that bilinguals suffer more from activation of the non-target language in the presence of cognates (e.g., Freeman et al., 2017). In the present experiment, nonword stimuli were used that have no connection to the English lexicon and were therefore unlikely to activate the non-target language. The question arises of whether this non-target language activation is also triggered by parallels on the sublexical level. If listeners hear a consonant cluster that belongs to both phonotactic systems, are they more likely to activate the non-target language than when they hear a cluster that is only legal in the target language? In such a case, English phonotactics would be activated and the influence of German frequencies should be reduced. The interaction between English and German phonotactics is in accordance with this idea. However, in that case, a main effect of English frequencies should also emerge due to the activation of English phonotactics. This is not seen in the data.

With respect to models of speech perception, the language specificity effect can be accounted for by considering frequency as a biasing factor, not as an inherent characteristic of connections between nodes, just as is assumed in the NAM. Here, the word decision units monitor higher-level lexical information, such as frequency, in order to bias activation levels.

6.7.7. Summary

Summing up, L2 listeners can be said to have more problems with acoustic interpretation because they are not as familiar with the specific phonetic realisations of consonants, the cues that can be used to differentiate between them, and language-specific inter-consonantal timing patterns. They therefore make more use of top-down knowledge about the structural characteristics of the target language, as can be seen from the stronger effect of German cluster frequencies. This effect is, however, mediated by the English cluster frequencies. Their stronger reliance on top-down knowledge sometimes leads to hypercorrection. It is also clear that L2 listeners are not influenced by sonority sequencing any more than native listeners. There are indications, however, that if universal principles are based on a more fine-grained measure with a strong foundation in psycholinguistic research—as NAD—then L2 listeners may display stronger effects than L1 listeners.

6.7.8. Conclusions and future directions

The analysis of the perception data corroborates the crucial role of acoustics and perceptual salience in speech perception. Overall, low error probabilities prevail for sC clusters and higher error probabilities for the other clusters. An exception to this pattern is /ts/, which has a very low error rate but could be argued to not be a true consonant cluster, but a single phoneme. Further research is needed to shed light on the role of /t͡s/ and its potentially special status. Accurate acoustic measurements are necessary to compare /t͡s/ to /ks/ and /ps/ in order to investigate whether acoustic differences determine its superior recognisability as a result of its affricate status.

Once the acoustic effect is controlled for, a significant frequency effect can be seen as well. Since L2 listeners are not as experienced with the exact phonetic realisations of phonemes, they rely on expectations of syllable structure even more heavily than L1 listeners. However, for L2 listeners, the effect of target language frequencies is modulated by L1 frequencies.

Taken together, even though our speech processing system is determined by our native language and this influence can be seen in the present data, it can be tuned towards using the kind of information from the L2 that is helpful for L2 listening to a large degree. Furthermore, the present study elucidates that structures that adhere to the SSP do not benefit from facilitation in perception, which is also demonstrated for L2 users, so usage-based factors are more relevant for perception than the SSP, a universal principle. In contrast, for NAD, which is more refined and has a stronger foundation in psychoacoustics, there was a significant effect in the L2 group only. This calls for further research on the role of other, more psychologically realistic, universal principles than the SSP; such as consonant sequencing based on NAD.

It must be kept in mind that, in the present study, the two phonotactic systems involved are very similar: structurally, English only differs from German in disallowing initial stop–sibilant and stop–nasal clusters (and allowing consonant–glide clusters). It would be very interesting to compare the present results to data from L2 listeners whose L1 differs more from the target language phonotactically (i.e., is more or less restrictive). In the past, such studies have hardly ever taken frequency distributions of phonotactic sequences into consideration. In order to further investigate the roles of L1 and L2 gradient phonotactics and their interrelationship in L2 listening, it also seems promising to test for a frequency effect of equivalent native clusters in structurally similar languages, for instance, an effect of English /sp/ frequency on German /ʃp/ perception. The low recognition rates for clusters that lack an L1 equivalent could be an indication that L2 structures gain additional support from L1 distributions of equivalent native structures. This would contribute greatly to our understanding of how the L1 and L2(s) are organised in bilinguals and how usage changes mental representations.

7. Sublexical speech production

Insights from phonological speech errors and a model

7.1. Introduction

Speech production is the process in which a speaker turns a mental concept into an audible utterance. Although the start and end points are the opposite of those in speech perception, the two cannot be said to be simple reversals of one another. There are a number of differences in processing steps and influencing factors. It is obvious that perception is constrained by acoustic factors, while production is influenced by articulatory ones.

In the following sections, the most important processes and influencing factors in word production (with respect to the study presented in the next chapter) will be summarised very briefly. Since processing accuracy is operationalised in terms of production errors, this chapter will discuss how speech errors serve as an insight into the processes and mechanisms of speech production. In addition, a connectionist model of speech production will be presented.

7.2. Stages and processes in speech production

Single word production can be divided into three major steps[1]: 1) conceptualisation 2) formulation, and 3) articulation. The step that is of concern here is formulation,[2] which can be subdivided into word selection and sound processing; sound processing can be further broken down into *phonological*, *phonetic encoding*, and *motor encoding*.

As anticipation errors show, the planning of an utterance occurs well ahead of its execution (sometimes several words, but anticipations usually occur within the same phase; cf. García-Albea et al., 1989), although both processes are executed in parallel.

The conceptualisation phase is thought to be preverbal. The speaker composes the message he or she wishes to convey in an abstract form, which is then passed on to the formulation stage in which lemmas (also abstract, but grammatically specified, representations) are selected according to the intended meaning. During phonological encoding, the phonemes that make up the selected lemma are retrieved and organised into a sequence with a specific stress pattern. Phonological processing proceeds sequentially, which means that the first phonemes in the sequence are prepared first. Support for this plausible idea comes from priming studies that showed earlier facilitation by primes that share initial phonemes with the target than primes that share later phonemes (A. S. Meyer & Schriefers, 1991). The output of phonological encoding—and the input to phonetic encoding—is a sequence of phonemes that is not context-specific, which means there is a lack of coarticulatory and allophonic specification (Buchwald, 2014). An alternative to the notion of phonological representations is provided by Articulatory Phonology, which does not feature phonemic repre-

[1] All information is based on Griffin and Ferreira (2006) unless stated otherwise.

[2] Even though sound processing is the stage relevant to the present study, articulation naturally also plays a role in the experiment laid out in Chapter 8 (as in any speech production experiment) and can add confounding effects that should be kept in mind or, ideally, controlled for.

sentations but rather articulatory gestures as the input for phonetic processing. It is during phonetic processing that context-specific phonetic details are brought forth (in Articulatory Phonology, the context-dependent temporal and spacial specifications) and phonetic representations created. Lexical characteristics can influence variation across phonetic representations. It has been shown, for example, that HF words are produced with shorter phoneme durations than LF words (Bell et al., 2009). This is even true for homophones, such as *time* and *thyme* (Gahl, 2008), but only applies to content words. The phonetic representations are then translated into motor programmes during motor encoding. As modelling of apraxic speech production data has shown, motor planning is not linear and sequential. Instead, articulatory gestures are hierarchically organised into syllables and metrical structures (Ziegler & Aichert, 2015). This organisation is demonstrated, for example, in the differential vulnerability of individual syllable constituents to apraxic speech errors, the facilitating effect of default metrical structure on production, and in variability in segmental articulation depending on metrical position (Ziegler & Aichert, 2015). The motor programmes are finally executed during articulation. Since motor processes are primarily concerned with “highly repetitive, not-too-varied and predictable relationships” (Keller, 1987; p. 128) between the input chain and coordinated articulatory movements, they can be automatised through learning. The processes of early planning stages, on the other hand, cannot be automatised because they are more variable and less predictable (Keller, 1987), although here, too, frequency and recency effects can be observed, which can be interpreted as a result of learning. The experiment on consonant cluster production in Chapter 8 can be taken as an examination of whether the stages in between word selection and motor execution can be automatised through the learning of frequent structures.

Although the stages of word production are generally agreed on, models diverge on whether they are strictly modular or interactive.

Does phonological encoding start only after lexeme selection is complete or can the phonological make-up of words influence their selection? In some models (so-called *cascaded activation models*; cf. Goldrick & Blumstein, 2006a), activated lexical representations can influence phonological encoding even when they are not selected. These models receive empirical support from studies that show shorter latencies for words that are phonologically related to synonyms of a word the speaker was preparing to say.

Similarly, there is some disagreement concerning the possibility of a feedback mechanism or a bidirectional flow of activation between levels. Some assume a monitoring system for output forms that filters non-lexical or phonotactically illegal forms, for example (e.g., Shattuck-Hufnagel & Klatt, 1979).

It is important to note that, although lemma and lexical selection are not required for the production of *pseudo*words, the lexical level can still interfere, for instance via the activation of neighbours (Vitevitch et al., 2014).

7.3. Factors in phonological, phonetic, and motor encoding

There are a number of factors with respect to the target item and the production process that can influence one or several encoding stages in speech production. Among them are lexical characteristics (such as lexical frequency and neighbourhood density), predictability, structural complexity, similarity between items, repetition, and position within the target item. In this section, the effects caused by each of these factors will in turn be summarised briefly.

Lexical frequency, as well as the predictability of a lexeme, do not only influence higher-level stages of speech processing, such as conceptualisation and lemma selection, but also lower-level processes; this is demonstrated, for instance, by the fact that phonetic variation depending on these variables. For example, less predictable words tend to be longer in duration and have a higher intensity than more predictable words (Lam & Watson, 2010).[3]

Lexical and sublexical frequencies have been shown to influence a number of processing stages, from phonological encoding (Laganaro, 2005) to phonetic (Laganaro & Alario, 2006) and motor processing (Ziegler & Aichert, 2015), although they cannot always be specified with certainty at a particular level (cf. Buchwald, 2014). The effects of frequency can manifest as differences in processing latencies (HF items take less time to process; Cholin & Levelt, 2009), differences in exact articulation (phonemes in HF words are reduced; Bell et al., 2009), or differences in production accuracy (more errors on LF items; Levitt & Healy, 1985). A detailed discussion of frequency effects on different linguistic units will follow in the next chapter.

Phonological similarity between items probably affects several subprocesses of sound processing. Its role is twofold, providing either support or competition, depending on which items exhibit similarity and how similarity is defined. On the one hand, words from dense neighbourhoods (i.e., words that are phonologically similar to many words in the lexicon in terms of shared segments) are produced faster (Vitevitch, 2002), are less prone to phonological substitutions and cause fewer tip-of-the-tongue (TOT) states (see also Section 8.2.3). Note that this effect of phonological neighbourhoods contrasts with that of semantic neighbourhoods, which have an inhibitory effect. On the other hand, similar sounding words can have an inhibitory effect, namely when they stand in syntagmatic opposition in a sequence. This property is utilised in

[3] Note, though, that the difference in duration between predictable and less predictable lexemes cannot be solely attributed to processing differences. Words that usually appear in predictable contexts are reduced in duration even when they are produced in a less predictable context, which suggests that the reduction is stored in the lexical representation (Seyfarth, 2014).

tongue twisters. For example, a shared onset in two words produced in succession slows production (Sevald & Dell, 1994). Sevald and Dell (1994) call this form of phonological competition *sequential cuing* and explain it in terms of activation spread over time: “If activation initially spreads to the onset, then to the vowel and then the final consonant, shared sounds produce competition between sounds that follow the repeated ones” (Sevald and Dell, 1994; p. 110). This is also true if the word with overlapping sounds occurred in the previous trial.

Similarity between phonemes in terms of shared features also creates competition. Pouplier (2008; p. 114) states that “the more two elements have in common, the more likely they are to interact in an error”. This competition on the featural level shows in gradient speech errors that exhibit features of both the target and competitor phoneme (Pouplier & Goldstein, 2010) and in a tendency for speech errors to involve featurally similar phonemes (Fromkin, 1971). Some accounts hold that the effect of featural similarity is particularly strong at the beginning of a word (Frisch, 2000): repeating words with similar onsets is significantly more difficult than repeating those with dissimilar onsets. Others have found equally strong effects for onset and coda similarity (Mooshammer et al., 2015).

Unsurprisingly, repeated elements become more easily accessible: they are repeated more accurately (Page & Norris, 2009), are produced with less phonetic prominence (Lam & Watson, 2010), and become reduced in duration (Bell et al., 2009). The reduction effect was not found in word list productions, however, which can be interpreted as evidence that the reduction effect functions at a level above phonological encoding in speech processing, for example, message planning (Lam & Watson, 2010). Nonetheless, they also create a competitive environment for the elements that surround them (as they are part of similar larger units, cf. the similarity effect described above) and thereby increase the error probability of such elements.

It has also been found that the position of a segment in a string of phonemes plays a role in its processing: onset consonants show more distinct articulation patterns than coda consonants, for example, tighter constrictions (Krakow, 1999). On the other hand, word onsets are more frequently involved in speech errors than elements that appear later in the word, including onsets of later-appearing syllables (e.g., Nooteboom & Quené, 2015). Dell (1986) attributes the word onset effect to higher activation levels at the onset, which allow competing onset consonants to more efficiently force their way into the sequence. However, no analogous effect could be observed for nonwords (Wilshire, 1998).

7.4. Speech errors as a window to speech production mechanisms

Speech errors, or *slips of the tongue*[4], are "unintended, nonhabitual deviation[s] from a speech plan." (Dell, 1986; p. 284). Using speech errors to infer aspects of the underlying speech production system has a long tradition, which goes back as far as Meringer and Mayer (1895) in the late nineteenth century. As Boomer and Laver (1989; p. 2) explain: "The general strategy is that of inferring relevant properties of an unobservable system on the basis of its output characteristics." In experimental speech production studies, errors provide a particularly valuable insight because they—being unintended—reveal "the production system that is uncontaminated by explicit knowledge" (Warker and Dell, 2006; p. 387). For many purposes, they therefore represent an advantage over methods like goodness ratings.

Speech errors show a remarkable level of regularity, which resulted in even the earliest speech error researchers to conclude: "[S]ie müssen

[4] In the literature, the terms *speech error*, *slip of the tongue*, and *lapse* are used to denote the same phenomenon. They will also be used interchangeably here.

durch konstante psychische Kräfte bedingt sein und so werden sie zu einem Untersuchungsgebiet für Naturforscher und Sprachforscher, die von ihnen Licht für den psychischen Sprechmechanismus erwarten dürfen." [They must be conditioned by constant mental forces and thus become a field of study for naturalists and linguists, who can expect them to shed light on the mental mechanism of speech.] (Meringer & Mayer, 1895; p. 9). Almost a century later, Fromkin (1973; p. 112) noted that slips of the tongue are "nonrandom and predictable" in the sense that the regularity allows researchers to predict the kinds of errors that can occur and those that cannot. This also implies that any given model of speech production that lays claim to reality and completeness must be able to account for all observed speech errors.

Before going into detail about the inferences that have been drawn from speech errors and their regularities, the following paragraphs will provide a quick classification of speech errors. In general, a distinction is made between *contextual errors*, which have their source within the utterance, and *noncontextual errors*, which do not. Contextual errors are subdivided according to the temporal relationship between the source and the error. In *anticipations*, the error element is produced before the source (i.e., anticipated), in *perseverations*, it is produced after the source (i.e., the source element perseveres in the production). In *exchanges*, two elements swap places. Normally, anticipations outnumber perseverations and exchanges (this is what Dell et al. (1997) call a "good" error pattern), but the number of perseverations increases as a function of the difficulty of the target utterance (Nooteboom, 1973). Noncontextual errors include *substitutions* of a linguistic unit by one that is not part of the utterance and *shift errors*, in which an element moves to a different place in the utterance.

Slips of the tongue occur on all linguistic levels, which has been taken as evidence for the psychological reality of various linguistic units and levels: lexical, syntactic, morphological, segmental, and even subsegmental (Frisch & Wright, 2002). For example, if two phonemes are exchanged in a slip like *heft lemisphere* for *left hemisphere*, it is assumed that phonemes are psychologically real units in speech production. Interestingly, not only phonemes and larger units can participate in a speech error like that but also phonological features, as in the exchange error *glear plue sky* for *clear blue sky* (both examples above are taken from Fromkin, 1973). This shows that features, too, have some degree of psychological reality, although Shattuck-Hufnagel and Klatt (1979; p. 41) note that phonological errors "almost always involve the movement of unitary segments and not the movement of component distinctive features" (cf. Nooteboom, 1973, for a similar observation). This dissertation is only concerned with phonological errors, which are considered the most common kind of speech error by many researchers (e.g., Boomer & Laver, 1989; Nooteboom, 1973).

Not only do slips of the tongue serve as evidence for linguistic units but also they are also indicative of a hierarchical structure of and within units. For example, exchange errors usually involve the same syllable and metric positions in both words, which shows that syllables are not indivisible units but have an internal structure. Another indication as to the internal structure of syllables comes from proportions of certain errors: while syllable onsets are often separately involved in an error, the vowel and coda tend to move or be substituted together (Stemberger, 1993), which shows their high internal cohesiveness as a rime. The error rate further depends on which class the final consonant belongs to. Most errors occur on stops, fewer on nasals and the least on liquids and glides (Stemberger, 1993). Thus, even within the rime, different levels of internal cohesiveness exist, which could be attributed to sonority sequencing. Similarly, it is conceivable that onsets consisting of several phonemes exhibit differing degrees of cohesiveness. This will be further discussed in the next chapter.

The fact that speech errors lead to phonotactically legal output in the vast majority of cases has been seen as evidence for the existence of phonotactic constraints (Dell, 1986; Fromkin, 1971), although its empirical foundations have been challenged recently (see also Section 8.2.1 in the next chapter).

With regard to the mechanisms of the speech production system, speech errors can provide insight into the modularity or interactivity, as well as the timing and order, of subprocesses. For instance, the fact that the indefinite article adapts to a new noun in noun ordering errors of the type *a courage of example* for *an example of courage* (Fromkin, 1973) demonstrates that the article is only spelt out after the reordering of the nouns occurred. Likewise, in phonological errors, the phonetic realisation of a phoneme often adapts to the phoneme's new environment. This can be seen in the aspiration of prevocalic stops after the deletion of an initial /s/, for example.

The fact that exchange errors make up 5% of phonemic errors but 20% of higher-order unit errors has been interpreted as an indication that memory decay rates of higher-order units are slower than those of phonemes: they must be kept in mind for a longer amount of time in order to complete the exchange. The distance between higher-order units involved in an error is also greater than between lower-order units. Nooteboom (1973; p. 154) concluded that "immediate memory in speech production is not a fixed amount of speech forms, but differs for the various hierarchical levels of linguistic units".

In general, speech errors have been taken as indications for the mechanisms of activation and competition in speech production models. Boomer and Laver (1989; p. 4) note: "In the bulk of our examples a plausible origin for the intrusion can be found in the immediate environment of the slip." From this it can be concluded that the elements to be used in the planned utterance become activated during the planning process and compete with one another. In some cases, this competition process can lead to interferences.

Evidence for feature-based competition between phonemes can be found in the articulatory variability of speech errors. If two phonemes in a tongue twister differ only in voicing, the VOTs in their production vary the most; if they differ only in place of articulation, the tongue-palate contact varies the most. If they differ in more than one feature, the variability of each of the features is smaller because the competition between them is not as strong (McMillan, 2008).

Slips of the tongue also serve as windows to control mechanisms in speech production. Both in error collections and in experimentally induced speech errors (Motley & Baars, 1975), a lexicality bias (a bias to produce an existing word) has been found. Particularly in the early days of speech error research, this was taken as an indication for an output monitor that prevents nonwords planned as the result of an error from being produced. In interactive models, the effect can be derived by feedback from lexeme units. One such model will be presented in the next section.

7.5. The Spreading-Activation Theory of Retrieval

One of the earliest and most influential network models of speech production is the Spreading Activation Theory of Retrieval in Sentence Production (Dell, 1986). It models speech production from syntactic planning via lexeme selection and morphological encoding to phonological encoding, with the same principles of spreading activation operating at all levels. Here, only the general mechanisms and the part of the model that is relevant to the present study, namely phonological encoding, will be described.

Although the theory is quite old and much empirically-inspired progress has been made since its postulation,[5] it still captures the speech production processes at work and the phenomena investigated

[5] For example, its localist architecture has largely given way to parallel distributed networks in more recent models.

in the experiment in Chapter 8 better than any other model I am aware of.

The Spreading Activation Theory combines a connectionist network architecture and a spreading activation retrieval mechanism with linguistic assumptions regarding units and rules. Like most theories of speech production, it assumes that producing an utterance involves the construction of internal representations—one for each linguistic level—that are built up before overt behaviour. The syntactic, morphological, and phonological representations are the results of syntactic, morphological, and phonological encoding, respectively. The separation of different linguistic levels, each with its own set of generative rules defining the "combinatorial possibilities" (Dell, 1986; p. 286) to construct the respective representations, is essential to the theory. The construction of representations at several levels occurs simultaneously and the model includes the possibility of interaction between them via feedback and cascaded activation. The generative character of the theory, particularly the strict distinction made between items in the lexicon and generative rules that work over these items, stands in contrast to usage-based theory in which such a distinction is not made. In the latter, chunks of various sizes are stored in the mental lexicon instead. It is possible that this discrepancy between linguistic traditions does not come to play at the sublexical level investigated in the present study. It will be interesting, however, to test this theory against experimental data that were collected with the aim of addressing a usage-based question.

As the name suggests, the spreading of activation from one node in the network to the next is the main principle of the model. The basic idea is that the spread of activation from nodes at a higher processing level activates nodes that may be used in the next lower representation. When a node becomes activated, it sends some of its activation to all nodes that are connected to it. This happens at a specific spreading rate *p*, which is independent of the speaking rate. Since all connections in the model are bidirectional, the higher-level node receives positive feedback from the lower-level nodes it sent activation to. Hence a given node receives activation both from top-down and bottom-up connections. All of this influx of activation adds to the node's current activation level. At some point, the node with the highest activation level is selected and used for the representation currently being constructed. When a node gets selected, its activation level is set to zero. However, because of its activated neighbours, which still send activation to it, it rebounds from zero. Over time then, it decays exponentially towards zero at a decay rate *q*. The model does not feature active inhibition between nodes, nor does it include any thresholds or saturation points; all activation dynamics are derived from excitatory connections and the principles of *spreading* (a node's property of sending out a proportion of its activation to all nodes it is connected to), *summation* (the sum of incoming activation to a node, which adds to that node's activation level), and *passive decay* (the gradual fading away) of activation. The model assumes *quantised time*, that is, discrete time steps. This means that all nodes' activation levels at each time step t_i can be calculated based on the activation levels at the previous time step t_{i-1}. The general formula for calculating the activation level A of a node *j* at a given time step t_i is given in (7.1).

$$A(j, t_i) = [A(j, t_{i-1}) + \sum_{k=1}^{n} p_k A(c_k, t_{i-1})](1 - q) \tag{7.1}$$

In the formula, c_k stands for the nodes c_1, c_2, c_3, ...c_k which are directly connected to *j* and, as mentioned above, *p* denotes the spreading rate and *q* the decay rate. Hence *j*'s activation level at t_i results from the level at t_{i-1}, the sum of incoming activation from all nodes it connects to, and the decay rate. Note that activation is a variable that can only take on positive values.

These are the general workings of the model as defined by its network architecture and principles, but there are some important limitations taken from linguistic theory as to which nodes can be used to build a representation. The model incorporates an idea developed in earlier (symbolic) speech production models (e.g., Garrett, 1980; MacKay, 1972) in which a frame consisting of slots that are specified for elements of a certain class and a stock of items that can be placed into these slots. First the frame is created. Via insertion rules—rules that specify what kind of items can be inserted into each slot—the slots are then filled with items from the lexicon, which are labelled for their category membership. As a node might need to be inserted into a frame several times (which is the case if it is repeated in the utterance being planned), it can receive more than one category tag. It can then be reused during production. In the frame for a syntactic representation, insertion rules would specify a sequence of word-class-specific slots, whereas the phonological frame would indicate the CV structure of a word (with consonants being specified for onset or coda position). The insertion rules select items strictly within the required category, so that they can only compete with items from the same category. As will be discussed later, this is the reason why, even in speech errors, vowels are never substituted for consonants in the model and vice versa. However, the rules under operation are also relatively general in nature. For example, the phonological rule taken from linguistic theory simply defines 'Syl → Onset Nucleus Coda'. It therefore only restricts items to be inserted into the slots with regard to their syllable position tag, thereby preventing a consonant tagged for coda position to be inserted into the onset slot. The model is not sensitive to language-specific phonotactic rules and is therefore capable of producing phonotactically illegal sequences. This gives the system the capacity to model both the productivity of the sound system and phonological speech errors (cf. Dell, 1986; p. 296). Nonetheless, for reasons related to the network and spreading activation dynamics, the system rarely produces phonotactically illegal sequences, as simulations revealed.

A representation at a given linguistic level can be thought of as a "collection of order tags that are attached to nodes in the lexical network, dictating the contents of the representation and their order" (Dell, 1986; p. 286). Which of the appropriately-labelled items gets inserted into a slot is determined by a decision rule based on the activation levels of the nodes (i.e., lower-level items). Specifically, the steps in the creation of a representation are the following:

1) activation of the current node[6], that is, increasing its activation level by an arbitrary amount of so-called *signalling activation*

2) spreading of activation from the current node to the nodes it is connected to; at the same time, the rules operating at the lower level (which are specific for that level) construct the frame for the emerging new representation; these rules are responsible for most of the ordering of the frame slots

3) filling of the frame slots according to insertion rules: of all items with the appropriate tag for a slot, the one with the highest activation level is selected

4) reduction of the selected node's activation level to zero (necessary to prevent it from being selected again)

5) changing the higher-level representation's current node.

Consider an example from the process of phonological encoding. Phonological encoding takes its starting point at an activated morphological representation that is to be mapped onto a phonological representation. The network needed for this mapping consists of nodes

[6] The *current node* is an important concept in the theory. It refers to the node of a higher-level representation that is currently being translated into the corresponding items at the next lower level. At the beginning of the process, it is the node of the higher-level representation that is labelled as first.

for morphemes, syllables, syllable constituents (i.e., consonant clusters and rimes), phonemes, and phonological features. The nodes for consonants and consonant clusters, as well as their features, are tagged for their position in the syllable: onset or coda. The phoneme level also includes a null element, which is used in place of a consonant (or consonant cluster) in open and naked syllables; this can cause some complications, as will be shown later. An example of the network structure can be found in Figure 7.1.

For a given morpheme sequence to be translated into phonological representations, the process begins with an assignment of the *current node* status to the first morpheme in the sequence; for example, the German morpheme *Blatt* in the sequence *Blattschneiderameise*. As the current node, it receives signalling activation (100 arbitrary units). The next morpheme in the sequence (i.e., the one tagged as second) also receives some amount of activation, known as anticipatory activation, which is of a smaller value (50 units). This anticipatory activation models the continued processing at the morphological level, where the upcoming units are processed in parallel with *Blatt* at the phonological level. From the current morpheme node, activation spreads downward to the syllable node /blat/ and on to the cluster node /bl/, the rime node /at/, the phoneme nodes /b/, /l/, /a/, and /t/, and the feature nodes connected to these phoneme nodes (e.g., [voiced], [stop], [bilabial], [alveolar], [liquid], etc.[7]).

[7] The theory only distinguishes phonological features on the three dimensions of manner of articulation, place of articulation, and voicing.

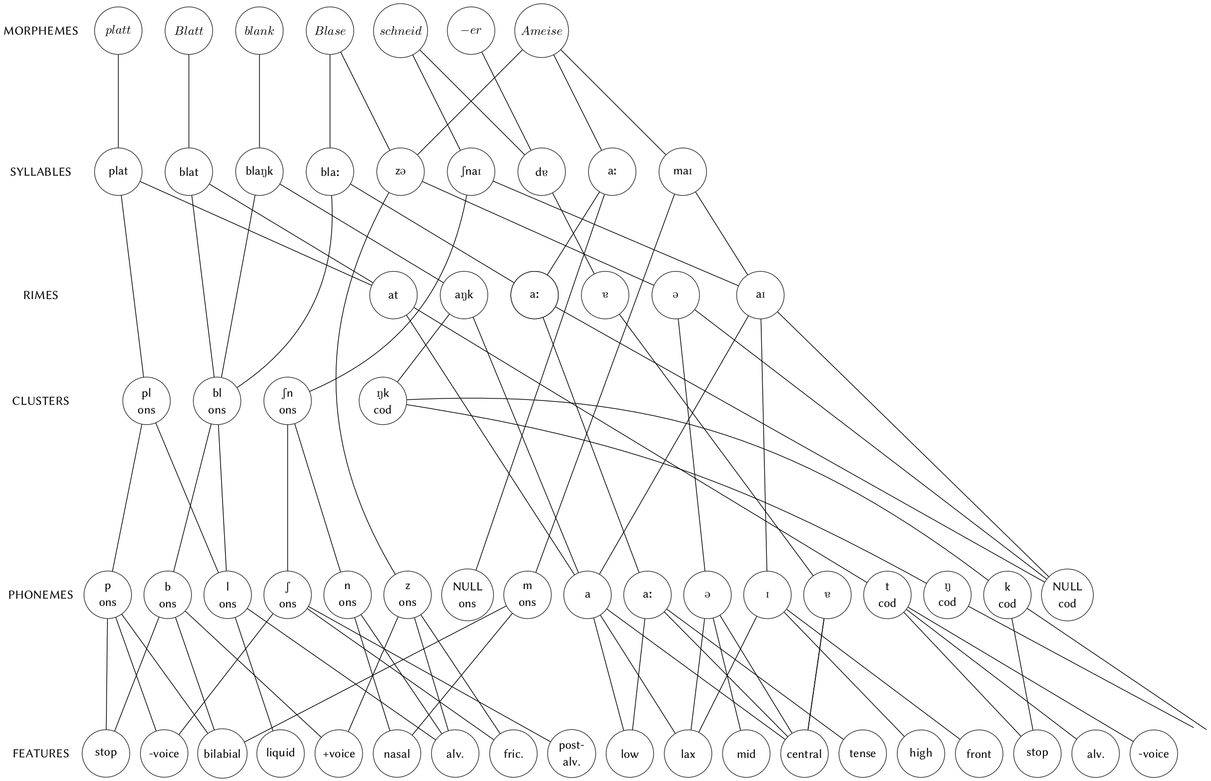


Figure 7.1.: A section of the Spreading Activation Theory network (adapted from Dell (1986)), not all nodes are shown for better readability

Activation spreads at a constant rate *p* through all downward connections and with a constant rate which is a fraction of *p* through all upward connections. Note that not only the nodes mentioned above become activated but also other nodes in the network. This is, on the one hand, partly due to anticipatory activation from *schneid-* being processed simultaneously at the morphological level and partly due to their connections in the lexicon and bottom-up feedback (e.g., the onset phoneme /d/ receives activation from the feature nodes [voiced] and [stop]). On the other hand, it is caused by noise coming from other sources (e.g., background activation coming from inferences that were made during the semantic planning of the sentence, cf. Dell, 1986; p. 291). The other nodes' activation levels are usually relatively low, though. While activation spreads from the morpheme node, the frame for the phonological representation is constructed as $\mathrm{CC_{ons}VC_{cod}}$.

After *r* time steps (with *r* representing the speaking rate), the most highly activated node for each slot in the frame is selected. In error-free production, the syllable constructed would consist of the onset cluster /bl/, the nucleus /a/, and the coda consonant /t/, which is the output for the phonological representation for that part of the utterance; their activation levels would accordingly be set to zero afterwards (*postselection negative feedback*). The *current node* status at the morphological level would then be assigned to *schneid-*, whose activation level would be increased by the signalling activation. This would lead to activation spreading to the syllable node /ʃnaɪ̯/ and further down in the network, and the process would repeat itself.

The Spreading-Activation theory is meant to model normal speech production, but it was developed on the basis of analyses of speech errors and, crucially, it can account for the patterns found in speech error data as well. Slips of the tongue are assumed to result from the co-occurrence of several units in a buffer that stores advance planning of higher processing levels. Processing problems can arise due to the synchronicity of representations constructed at several levels and the feedback from lower to higher levels. When a node that is not the current node at the higher level any more receives sufficient activation in the form of feedback from the lower-level nodes it is connected to (because they are the current nodes at the lower level) and belongs to the same category as the current node, then the insertion rule might err. It may then insert that node into the slot instead of the correct one. Dell calls that a problem of "confusion between levels". This mistake is only possible because items can be reused (see above). The same problems arise when an upcoming unit becomes too highly activated by anticipatory activation and the resulting activation spread. These two scenarios can create contextual slips of the tongue. In addition, nodes in the network that are not part of the utterance being planned can become activated due to their connections, creating potential for non-contextual errors. All important predictions that the theory makes concerning speech errors are derived from spreading activation. In the following paragraphs, the main factors influencing the probability of production errors according to the Spreading-Activation theory will be discussed briefly.

Most obviously from an empirical perspective, speaking rate influences error probability. The faster a person speaks, the more often he or she tends to slip up. In the Spreading-Activation theory, speaking rate is represented by the parameter *r*, which is defined as the number of time steps per syllable. It determines the rate at which the slots of the frame become available for filling. When *r* has a high value, more time passes between the encoding of two syllables than when it has a low value. Since the spreading and decay rates in the model (i.e., *p* and *q*, respectively) are independent of the speaking rate, cases can arise in which the speaking rate does not allow for the spreading and decay of activation necessary for the selection of the correct item. Therefore, the wrong nodes might be the most highly activated at the time of selection, and speech errors can occur. The activation of a previously activated element might not have enough time to decay and/or the spreading rate might not be high enough for the target nodes of the current representation to be sufficiently activated. From this it follows that perseveratory errors should dominate at fast speaking rates, which is also what is commonly observed empirically (e.g., Dell et al., 1997). The specific interactions between speaking rate and other factors will be identified below.

Another factor that influences speech errors is the distance between items: items involved in an error, for example the participants in an exchange error, tend to be relatively close together. This holds for words as well as phonemes and linguistic units at an intermediary level. In the framework of the Spreading-Activation theory, this effect can be explained in terms of activation levels. The most highly activated competitors for an item are those that have just been selected (because their activation has not had enough time to decay yet) or the ones that are about to be selected (because of anticipatory activation, i.e., parallel processing at a different level).

As discussed in Section 7.4, similarity is an important factor in the creation of errors. This includes similarity between phonemes and similarity between syllables due to shared phonemes (the repeated phoneme effect). The Spreading-Activation theory can account for both in terms of network structure and spreading activation. Similarity between two phonemes is generally defined in terms of shared features. For example, the phonemes /b/ and /p/ are very in similar in that they share all features except [voice]. According to the Spreading-Activation theory, the features activated by the activation flow from the phoneme nodes will send feedback up to the phoneme level, thereby also activating similar phonemes to a certain degree. During the processing of /b/, activation will therefore spread to the feature nodes [stop] and [bilabial], which would subsequently send bottom-up activation not only to [b] but also to [p]. However, when the speaking rate is high, there might not be enough time for the feature nodes to send activation back to phoneme nodes and activate phonemes similar to the target. Consequently, similarity effects will be considerably smaller at fast speaking rates. The case is analogous for repeated phonemes one level higher: syllables that share a phoneme are indirectly connected via the node for that phoneme. The phoneme node will receive a higher amount of activation because it comes from several syllable nodes and will send feedback to all syllable nodes it is connected to. For that reason, it is "more difficult to keep sounds from the wrong syllable from getting a high activation" (Dell, 1986; p. 301). Coming back to the example above, the repetition of /aɪ̯/ in the syllables /ʃnaɪ̯/ and /maɪ̯/ of *Blattschneiderameise* presents a potential processing problem. During phonological encoding, the phoneme will activate all syllables it is contained in. Since /ʃmaɪ̯/ is one of them and receives feedback from all of the activated phonemes, an anticipatory error is more likely to occur than in contexts without phoneme repetition (and, in this case, it is further facilitated by the featural similarity between /m/ and /n/). Thus, while the repeated phoneme itself has an increased chance of being produced, the surrounding sounds are more error-prone.

The Spreading-Activation theory also models a number of outcome biases observed in human speech errors. Among them are the lexicality effect, the syllable bias, and frequency biases. By and large, they depend on feedback loops once again, as explained with respect to the other effects described above. These loops between higher- and lower-level nodes "cause[] the activation levels to adjust themselves so as to reflect the entire stored vocabulary" (Dell, 1986; p. 300) and its characteristics, including sublexical characteristics. For all of the biases, the speaking rate has to be sufficiently slow to allow for the feedback loops to develop. At higher speaking rates, there are fewer opportunities the feedback mechanisms to "repair" high activation rates of the "wrong" nodes. Therefore, there is stronger bias for lexical output, frequent phonemes, and frequent phoneme combinations at slower speaking rates.

The lexicality bias (i.e., the tendency for an error to create an existing word or morpheme rather than a nonword) follows from syllable nodes sending activation to morpheme nodes. Coming back to the example presented above, the activation from *Blatt* will spread downward through the network and upward again via feedback. As has been explained, feature nodes also activate the nodes of similar phonemes not present in the morpheme, so /b/ will also indirectly activate /p/ because of their similarity and /p/ will send activation to syllables it is contained in. As a result, the syllable /plat/ will be highly activated because it shares most of its phonemes with /blat/ and receives activation from two of p's features as well. As this syllable happens to be a lexeme, the morpheme node *platt* will give its activation level a further boost. Therefore, it has a higher chance of output than a syllable that does not constitute an existing morpheme/lexeme. The syllable bias (i.e., the tendency to create existing syllables) can be explained in a parallel fashion.

The same holds also for a phoneme frequency effect. A phoneme that is frequent (i.e., is part of many different syllables and morphemes) receives more activation from these nodes than a less frequent phoneme, so the principle of summation causes its activation level to be higher than that of less frequent phonemes. In the example above, /b/ would receive more activation than /p/ because it is more frequent as an onset in German (i.e., has more connections to different syllable nodes) than /p/. In addition, the onset cluster /bl/ is slightly more frequent than /pl/. This is what Dell (1986) calls *contingent frequency*: in the context of a following /l/, /b/ is more likely than /p/. The odds would be reversed in the context of a following /r/ because /pr/ is a more common onset than /br/.

In general, Spreading-Activation theory predicts phonological speech errors to "create units whose nodes have many connections to them" (Dell, 1986; p. 301) or, put differently, "[T]he model finds it easiest to encode strings of sounds that are likely to be correct—words, morphemes, and syllables that exist in its vocabulary, and frequent sounds and sound combinations" (Dell, 1986; p. 301).

This means that the frequency effect that is the object of investigation here follows naturally from the structure and working mechanisms of Spreading-Activation theory. From the output bias for HF sequences, it follows that LF sequences will be harder to produce.

The question arises as to how the model would deal with sonority preferences. It does not have an inbuilt device that would bias it towards outputting SSP-conforming syllables. One way sonority sequencing could be implemented is by varying connection weights between nodes. In the original model, all connections have the same strength (i.e., the spreading rate *p* has the same fixed value for all top-down connections and another fixed value for all bottom-up connections). However, Dell himself suggests that this parameter could be modified in order to make model performance more similar to human performance (cf. Dell, 1986; albeit on a different subject). Thus a sonority bias has the potential to be implemented. However, it does not follow from the model's structure and mechanisms, nor does it seem to be independently motivated.

Spreading-Activation theory has been computationally implemented (Dell, 1986, 1988) and successfully models a number of speech error phenomena from human speech production. Among them are the prevalence of phonological speech errors, the effects of speech rate, similarity, and the outcome biases described above, in addition to the syllable position effect and the most common error types (anticipations, perseverations and exchanges), as well as their proportions. Moreover, it generally produces phonotactically acceptable sequences (although it occasionally produces final syllables that are unacceptable in word-final position). Nonetheless, it does deviate from human behaviour in some respects. Unlike humans, it cannot produce syllable deletions and is not sensitive to the stress pattern of utterances, simply because stress is not implemented in the model. In human speech errors, exchanges occur between syllables with equal stress and more often on stressed syllables. It also diverges from the results obtained in experimental studies since it shows a deletion bias for consonants, while human speakers normally display an addition bias (Stemberger, 1991). This behaviour of the model is due to the use of null elements among the phoneme nodes, which have a relatively high frequency and thus can act as intruders more easily than other nodes. As a solution, Dell (1986) offers an alternative model implementation that does not involve null elements. In this alternative, the frames would be constructed under the guidance of the form-related nodes in the network. In the version described above, the phonological frame was simply constructed by the rule 'Syl → Onset Nucleus Coda', which was repeated as long as there were morphemes to be translated into phonological representations. In the alternative implementation, frame-building would be informed by the nodes in the network and their activation levels so that an open syllable would not have a coda slot and, consequently, there would be no need for null elements. In this case, addition and deletion errors would be a result of the wrong frame being built.

As the above discussion shows, Spreading-Activation theory has a few shortcomings and cannot be taken as a perfectly accurate model of human speech production in general or of phonological encoding specifically. It is also disadvantageous in that it stops at the point of the completed phonological representation and does not capture phonetic/motor encoding or articulation. However, for the processes and phenomena under investigation in this dissertation, it is a reasonable approximation. It could probably be substantially improved by allowing the connection weights between nodes to vary as a function of different characteristics of the units as represented by the nodes. Varying connection weights are a feature of virtually all recent models.

Finally, none of the models available seem to be able to model exactly the processes or the effects of the experiment described in the next chapter . For example, OSCAR (Vousden et al., 2000) and HOR-

ROR (H. D. Harris, 2002) model only production of CVC words (or, with a different parameter setting, words of any other *predefined* syllable length), WEAVER/WEAVER++ (Levelt et al., 1999; Roelofs, 1997) disregards frequencies of units smaller than the syllable, and the Spreading Activation theory's SRN successor (Dell et al., 1993b) does not produce contextual errors. A cascading-activation model, such as the one discussed by Goldrick and Blumstein (2006b), seems promising but has, to the best of my knowledge, not yet been described in detail or computationally implemented.

8. Experiment 3: Production of German consonant clusters

8.1. Introduction

What factors influence which linguistic items are hard or easy for our speech production system to process? Speech errors are generally taken as an indication of competing phonological representations or speech plans (Baars, 1980; Dell, 1986; McMillan & Corley, 2010), which have been activated simultaneously. The error then results from one of the competitors of the target item erroneously winning the competition for production due to higher activation than the other items. What exactly makes some candidates “stronger”, that is, more highly activated, therefore enabling them to win the competition more often is an intricate matter that is still open for debate. A number of factors that make a candidate (phoneme, syllable, or word) stronger have been identified, including its frequency of occurrence. However, evidence for the role of frequency on various linguistic levels is inconclusive. Most studies suggest that speech errors tend to replace LF items with HF items (Levitt & Healy, 1985; Motley & Baars, 1975), but some have found either no correlation (Croot & Rastle, 2004) or an effect in the opposite direction (Santiago et al., 2007). Another line of research indicates that, at the sublexical level, it is primarily phonologically marked linguistic structures that fall victim to speech errors (Miozzo & Buchwald, 2013).

This study is aimed at testing the relative influence of consonant cluster frequencies and sonority sequencing on the production of pseudowords. In order to do so, the 16 test clusters used in the perception experiments were arranged in pairs of similar clusters that either do or do not differ in frequency/sonority. From these onset cluster pairs, pairs of monosyllabic pseudowords were created to be repeated in a tongue twister paradigm.

8.2. Previous research

In the following sections, an overview will be given of the literature on the roles of sublexical frequencies, neighbourhoods, and sonority hierarchies in speech production.

8.2.1. Phonotactics and frequency in speech production

Our speech production is efficaciously influenced by our phonotactic knowledge. It has been shown that children with developmental phonological error patterns, like fronting or stopping in consonants, exhibit such error patterns significantly less often if they would lead to the creation of illegal consonant clusters (Ott et al., 2006). For example, children who display fronting in singleton consonants in most cases (e.g., *Kuh* “cow” /ku:/ ⟶ [tu:]) do so considerably less in initial stop + /l/ clusters since this would result in the illegal sequence /*#tl/ (e.g., *Kleid* “dress” /klait/ ⟶ [*tlait], examples taken from Ott et al. (2006)).

It has also been noted early on in speech error research that slips of the tongue are “practically always” phonotactically legal (Wells, 1951; p. 86; see also Boomer & Laver, 1989). This has also been found for aphasic errors (e.g., Romani et al., 2011; Stenneken et al., 2005), which can be taken as an indication of how strongly phonotactics are coded into our speech production system, even in cases of malfunctions. However, using collection methods that are less prone to perception biases (cf. Cutler, 1981) than those traditionally used, it has recently been shown that phonotactic irregularity in speech errors of unimpaired speakers is considerably higher than previously assumed (Alderete & Tupper, 2018). The rate differs, however, depending on where in the syllable the error occurs. With respect to CC-onsets, for example, a substitution error on C1 produces phonotactically legal structures 81% of the time , which is not significantly different from chance, that is, what would result from substituting C1 with a random consonant. The speech error outcomes of C2 substitutions, on the other hand, are without exception phonotactically legal, which is significantly above the chance level value of 83% (Alderete & Tupper, 2018). Thus, although the originally observed high rate of adherence was partly based on perception biases and the actual rate is probably not as high as first thought (Alderete & Tupper, 2018), a strong tendency remains for segmental speech errors, at least in certain positions, to adhere to the phonotactic rules of the language spoken.

Moreover, experimental studies demonstrated that phonotactic constraints can be learned during the course of an experiment and that this newly acquired implicit knowledge also affects the shape of subsequent speech errors (Dell et al., 2000; see also Goldrick, 2004; Warker et al., 2009). During rapid repetition of nonce words that were consistent concerning the restriction of individual phonemes to certain syllable positions (e.g., /f/ can only appear in onset position and /s/ only in coda position), far more than 90% of participants' error productions respected the "experiment-specific phonotactics". Language-wide (i.e., English) phonotactic constraints, such as /h/ being restricted to the onset and /ŋ/ to the coda, on the other hand, were never violated. The finding that language users are able to learn phonotactic constraints over a relatively short period of time and are influenced by this newly acquired knowledge in their subsequent productions shows how sensitive speakers are to phonotactic rules, even after brief exposure. This contrasts with Alderete and Tupper's (2018) conclusion that even adherence to language-wide phonotactic rules is not as strong as first thought.

More importantly for the present study, however, the implicit learning of phonotactic regularities is not limited to categorical phonotactic constraints, such as /f/ being restricted to onset position, but also applies to gradient phonotactics of the type "labiodental fricatives occur in onset position 75% of the time and in coda position 25% of the time" (Goldrick, 2004). In an experiment, categorical constraints were hardly ever violated in speech errors, whereas gradient ones showed a smaller effect that was still significantly different from baseline, as expected. The two kinds of constraints also interacted: Goldrick (2004) demonstrated that the presence of featurally similar phonemes in "prohibited" syllable positions leads to an attenuation of the segmental phonotactic effect.

The occurrence of gradient (i.e., distributional) phonotactic effects is not surprising since frequencies have long been known to influence speech production. The word frequency effect, an advantage for HF lexemes, was one of the earliest to be described (e.g., Oldfield & Wingfield, 1965) and has been replicated countless times (e.g., Andrews, 1992; Levelt & Wheeldon, 1994; Mooshammer et al., 2015).[1] Moreover, facilitative frequency effects in speech production have also been observed on other linguistic levels using a plethora of methods, which suggests that the enhanced processing of HF items is a general mechanism in speech production and is applicable to all linguistic units relevant in processing. According to usage-based linguistics, this is due to entrenchment. For example, Bybee (1999; p. 232) describes the atoms of lexemes as "a set of highly entrenched gestures and gestural configurations that are used and re-used". The more often a particular gesture is used, the more entrenched it becomes and the easier it is for it to be executed. A more general explanation, not restricted to speech gestures, comes

[1] Note, however, that for Spanish, a negative word frequency effect has been reported, namely HF words leading to more production errors, cf. Santiago et al., 2007.

from connectionist models in which HF units are more strongly activated.

Nevertheless, even though the majority of studies that investigate frequency effects at various linguistic levels find evidence for them and they are plausible from a theoretical perspective, there are also some studies that fail to find any such effects. In the following section, the literature on frequency effects at different sublexical levels will be briefly reviewed and their dependence on various experimental factors disentangled.

The evidence concerning phoneme frequency effects is mixed. In their seminal study on speech errors, Shattuck-Hufnagel and Klatt (1979) found a correlation between the frequency of a consonant and its participation in speech errors, but the consonants' involvement as targets and intruders was symmetrical. Other studies found a tendency for HF phonemes to replace LF phonemes in speech errors (Goldrick, 2002, 2004; Levitt & Healy, 1985; experiment-internal frequencies: Goldrick & Larson, 2008). Stemberger (1991) found a frequency effect in non-contextual errors, but an "anti-frequency effect" (Stemberger, 1991; see below for a short discussion) in contextual speech errors. Goldrick (2003) found a phoneme frequency effect but only with respect to generalised phoneme frequencies, that is, taking phoneme classes into consideration. Finally, Santiago et al. (2007) observed what they dubbed "the David effect" (after the biblical story): the tendency of LF phonemes to replace HF phonemes in speech errors. In addition, latency-based phoneme frequency effects were found: in a naming task, both high word and high phoneme frequencies significantly sped up reaction times (Mooshammer et al., 2015). In the same experiment, however, execution time was only influenced by word and syllable frequencies, not phoneme frequencies. All of these effects were furthermore limited to immediate naming; in a delayed naming task with the same materials, none of the frequency measures affected either reaction or execution time.

On the syllable level, research has focused mainly on reaction time experiments, although some error analysis studies exist as well. Many studies find shorter production latencies for HF as opposed to LF syllables in pseudowords (Bürki et al., 2015; Cholin et al., 2011; Cholin et al., 2006; Laganaro & Alario, 2006) as well as words (Levelt & Wheeldon, 1994; computer simulation: Wade et al., 2010).[2] In analogy to the description given above for the phoneme level, syllable frequency effects emerge in immediate, but not in delayed, naming (Bürki et al., 2015; unless an intervening task prevents the preparation of syllables, cf. Laganaro & Alario, 2006).

This quantitative processing advantage for HF syllables led Levelt and colleagues to postulate a *mental syllabary* in analogy to the mental lexicon (Levelt, 1992, 1993; Levelt & Wheeldon, 1994; see also Crompton, 1981; for an earlier suggestion of a similar concept). A considerable amount of research has since been dedicated to the investigation of this mental syllabary. The idea is that the motor programmes for HF syllables are stored as holistic units that can be accessed during speech production, which makes it more efficient. LF syllables, on the other hand, have to be assembled on-line from their component segments. This latter process is thought to take more time, which explains the syllable frequency effect often found in reaction times. The entries in the mental syllabary are thought to take the shape of "phonetic programs or gestural scores" (Levelt, 1993; p. 293) and, accordingly, the stage at which the mental syllabary is accessed is assumed to be phonetic encoding. The fact that syllable retrieval speeds up during immediate naming and naming after an articulatory suppressor task (an activity that is assumed to leave phonological encoding relatively intact) but not during delayed naming without intervening tasks supports the

[2] Note, however, that Croot and Rastle (2004) did not find a latency effect, Carreiras and Perea (2004) found an effect only of first-syllable frequency but not of second-syllable frequency, and Brendel et al. (2008) report an interaction with syllable complexity, but no main effect of syllable frequency.

view that the locus of the mental syllabary is phonetic encoding (see Laganaro & Alario, 2006; for a detailed discussion).

In contrast with the findings from reaction time experiments, neuroscientific studies often fail to find syllable frequency effects. In several fMRI studies with pseudoword reading tasks, no hemodynamic effects of syllable frequency were found, but only effects of syllable complexity (Brendel et al., 2008; Riecker et al., 2008); this is interpreted as evidence against a mental syllabary and the processing of "syllable-sized phonetic representations [...] as holistic units" (Riecker et al., 2008; p. 111). In a relatively similar fMRI study involving repetition of auditorily presented pseudowords, on the other hand, pseudowords with frequent first syllables produced lower activity in a number of brain regions than pseudowords with infrequent first syllables (Tremblay et al., 2016). In contrast, the frequency of the second syllable of the disyllabic stimuli did not show a hemodynamic effect. Likewise, a pseudoword completion study found significant differences in the electrophysiological patterns between HF and novel syllables from around 170 ms to 100 ms prior to articulation onset (Bürki et al., 2015). LF and novel syllables did not differ in terms of activation patterns. These results, in turn, support the notion of a mental syllabary that reduces processing load for HF syllables but does not provide an advantage for LF syllables over novel ones.

In terms of error rates, even studies that found reaction time or neurological effects often find no differences between HF and LF syllables (Cholin et al., 2011; Croot & Rastle, 2004; Laganaro & Alario, 2006; Levelt & Wheeldon, 1994; although see Staiger & Ziegler, 2008; for frequency-based differences in error rates in apraxic and healthy subjects). An explanation for this lack of a qualitative difference within the frame of the mental syllabary is that on-line assembly of syllables' phonetic code, though slower than accessing pre-compiled representations, is still sufficiently accurate. However, a minority of studies do find differences in error rates. For example, the hemodynamic difference for first-syllable frequency in Tremblay et al. (2016) was accompanied by a difference in error rates. Santiago et al. (2007) found not only accuracy differences as a function of word and phoneme frequency but also syllable frequency in their analysis of natural speech errors. Recall, however, that the syllable frequency effect in their study is *negative*; this means that both target and error syllables have a lower frequency than the average syllable, and, in some of their analyses, error syllables have a lower frequency than target syllables (although this seems to be due to a correlation with phoneme frequencies, which are the origin of the effect, cf. Santiago et al., 2007). These findings therefore contrast with the general pattern of results obtained with regard to syllable frequency effects and should thus be viewed with caution. More convincing results come from studies on impaired language production, where HF syllables have been found to be more error-resistant than LF syllables (e.g., Aichert & Ziegler, 2004).

An intermediary level between the phoneme and the syllable—which is the object of the present study—has not been examined as often as the phoneme or syllable levels. However, there are a few studies that investigate the frequency effects of diphones, most of which do find effects. For example, accuracy differences between LF and HF bigrams[3] have been observed both for delayed word naming (Andrews, 1992) and pseudoword repetition (Edwards et al., 2004; the effect being more pronounced in children than in adults), although Munson (2001) finds the accuracy effect in pseudoword repetition to be significant only for children but not for adults. Importantly, Edwards et al. (2004) observed that, in errors on LF sequences, all substitutions resulted in higher frequency sequences (with a significant frequency difference). The outcome sequences of errors on HF biphones, on the other hand, were of a higher frequency than the target only in 50% of cases. Moreover, a more abstract skeletal structure seems to influence the effect of bi-

[3] The term *bigram* will be used here to refer to both diphones and bigraphs.

gram frequencies: Edwards et al. (2004) report that CV sequences were generally produced more accurately than expected based on individual biphone frequencies alone.

In terms of latencies, bigram frequencies failed to show effects in standard and delayed naming (Andrews, 1992) and led to a non-significant effect in word repetition (Bose et al., 2007). What is often observed, however, is that bigram frequencies influence word and/or segment duration and the specific articulatory characteristics. In general, duration seems to be negatively correlated with biphone frequency (word duration: Bose et al., 2007; segment/biphone duration: Edwards et al., 2004; Munson, 2001; Pouplier, Marin, Hoole, et al., 2017), which indicates that HF structures are produced more fluently.[4] Pouplier, Marin, Hoole, et al. (2017) report that, at higher speech rates, Russian HF onset clusters are produced with more overlap (i.e., coarticulation), while the degree of coarticulation does not change with speech rate for LF clusters. LF sequences are also produced with greater variability (Munson, 2001).

All in all, bigram frequencies seem to cause effects primarily on exact articulation parameters such as segment duration and overlap and much less on reaction times, which are an indication of processing difficulty before articulation. They can have an effect on error rates under certain circumstances, but note that for example in Andrews (1992), effects of bigram frequencies were less prominent than those of word frequencies or neighbourhoods. It should be noted, however, that the cohesiveness of the biphones studied varies considerably, which probably contributes further to the diverging results regarding frequency effects. While Pouplier, Marin, Hoole, et al. (2017) studied onset clusters, which arguably have a high internal cohesiveness, the object of study for Munson (2001) was heterosyllabic consonant clusters. The latter are, naturally, thought to be less unit-like and therefore were not

[4] As with error rates, Munson only finds an effect of duration in children but not in adults.

expected to show strong frequency effects. In light of this, it is remarkable that Munson (2001) found effects of biphone frequency for three different measures (accuracy, segment duration, and segment variability), even though two of them were only significant in children.

In addition to phone, biphone, and syllable frequencies, phonotactic or transitional probabilities calculated over several segments up to the whole stimulus item have been studied as a more holistic frequency measure. Although a distinction should be made between phonotactic probabilities based on positional phone and/or biphone frequencies and transitional probabilities as the probability of one phoneme given another, they will be addressed together here. Both measures share the capability to capture the probability of a longer sequence (e.g., a syllable or a polysyllabic sequence) without referring to the frequency of that sequence itself. This means they can assign non-zero probabilities to unattested sequences. *Phonotactic probability* refers to the summed positional phoneme or biphone probability of a sequence of segments. This measure is extensively used by Michael Vitevitch and his collaborators (e.g., Vitevitch et al., 2004; Vitevitch & Luce, 1998, 2005) who calculate it from the sum of all positional segment probabilities and the sum of all positional biphone probabilities in the word or nonword under consideration. *Transitional probability*, on the other hand, denotes the probability of a segment given the preceding (forward transitional probability) or following (backward transitional probability) segment. This means it is calculated as the probability of the transition under consideration in relation to all other possible (i.e., legal) transitions from the reference segment.[5] Like phonotactic probability, it is most often calculated over the whole word or nonword. One of the first studies to investigate an effect of transitional probabilities was a

[5] Note that Edwards et al. (2004) and Munson (2001), discussed under *bigram frequencies* above, call their frequency measure transitional probability even though it does not involve the probability of a *transition* from one segment to the next.

SLIP[6] experiment by Motley and Baars (1975). The investigators observed that more errors were made in which the intruding phoneme is more probable (in terms of transitional phoneme probability as well as positional probability) than errors in which it is less probable than the competing phoneme. What was important for the creation of contextual errors was thus the probability of potential intruders in the environment: error probabilities increased if one of the phonemes chosen for production was more likely in an earlier position in the utterance. Motley and Baars (1975; p. 360) conclude that

> spoonerisms apparently are facilitated when one of the phonemes in a phoneme string destined for articulation enjoys a greater probability of occurrence in an earlier-than-intended context than does the phoneme originally intended for that context.

They interpret this as evidence that sensitivities to transitional probabilities are active in cognitive processing of language.

These sensitivities to transitional probabilities are not only manifested in speech error tendencies, however, but also in terms of production latencies. Reaction times were found to be shorter for high-probability words than for low-probability words in different naming tasks (Kawamoto & Kello, 1999; Vitevitch et al., 2004). This effect of phonotactic probability was more stable than that of neighbourhood density (Vitevitch et al., 2004), thus demonstrating the strong facilitative (relative to inhibitory) influence of phonological similarity in production tasks. The authors interpret this finding as an indication of the

[6] The SLIP (Spoonerisms of Laboratory Induced Predisposition) technique is an experimental technique in which participants are primed towards a recurring sequence of word onsets by reading sequences of words that begin with those onsets, for example *dove ball*, *deer back*, *dark bone* (cf. Nooteboom, 2005). After a number of word pairs following one onset pattern, a word pair in which the onsets are reversed is presented, for example *barn door*, which causes participants to produce spoonerisms of the kind *darn bore*.

dominance of sublexical, as opposed to lexical, representations during speech production. Likewise, repetition of nonwords of high phonotactic probability is faster than that of nonwords of low phonotactic probability (Vitevitch et al., 1997). However, the latter experiment involved speeded repetition, a method that involves both perception and production components, so that it cannot be concluded with certainty that the effect originates from production processes. While Vitevitch and colleagues test for phonotactic probabilities in general (based on positional phone and biphone probabilities, see above), Kawamoto and Kello (1999) discovered that it is specifically *backward* transitional probability that is the best predictor of production facilitation. In a reading task, they found a positive correlation between the reaction time and the number of possible first letters given the second letter of a word. Conversely, the greater the probability of the first letter (given the second), the shorter the reading latencies. In the same experiment, they also found differences in error rates for high- vs. low-transitional-probability words, which were found to be significant by participant but not by item. Interestingly, they observed shorter latencies for onset clusters (significant for /s/–stop clusters and /sm/, non-significant for stop–liquid clusters and /sl/) than for simple onsets and partially lower error rates for onset clusters than for simple onsets. These observations—enhanced processing of a consonant cluster relative to a single consonant in spite of its length—support the assumption of the entrenchment of consonant clusters.

In spite of this strong evidence for a facilitating effect of frequencies at various linguistic levels, Stemberger (1991, 2004), in particular, has described what he called "anti-frequency effects" on the phoneme level: "[H]igh-frequency defaults[7] show greater error rates when competing with low-frequency nondefaults" (Stemberger, 2004; p. 415). For example, the HF [+anterior] phoneme /s/ is more often replaced by

[7] Stemberger uses the terms *default* and *nondefault* in the sense of universal markedness hierarchies on the segmental and featural levels.

the LF [−anterior] phoneme /ʃ/ in contextual speech errors than the reverse. Likewise, singleton consonants are often replaced by consonant clusters that are both structurally more marked and have a lower frequency. Stemberger (1991) traces these tendencies back to a single underlying principle, which he calls the *Addition Bias*. This explanation is based on the notion of underspecified representations, which means that default features like [+anterior] or [−voiced] are not specified in the phonological representations of words but are added during articulation if no other value for that feature is specified. If one assumes that no [anterior] value is specified for /s/ and that an Addition Bias is active in speech errors, then the tendency for /ʃ/ to replace /s/ is a natural result since the value [−anterior] is erroneously added to the underspecified representation of the sibilant. In this way, the tendency for nondefault phonemes to replace default ones can easily be explained in terms of the same principle that causes phoneme addition errors,[8] both of which seem to be at odds with general frequency effects.

Interestingly, in Stemberger's (1991) study, frequency effects on the phoneme level became apparent at the exact point when the Addition Bias did not apply because both competing phonemes were specified for a given feature (e.g., when a labial competed with a velar, both nondefault phonemes specified for place of articulation) and no feature value could thus be added; in such cases, there was a bias towards the more frequent phoneme. This shows that underlyingly present frequency effects can be hidden by stronger effects in the opposite direction, such as the Addition Bias. It is therefore important to identify these opposing effects and control for them in order to identify potentially weaker frequency effects. According to Stemberger (1991), frequency effects are visible in non-contextual errors, while anti-frequency effects arise in contextual errors. This distinction cer-

[8] For example, Stemberger (2004; p. 416) notes that, in SLIP experiments, "singleton consonants are prone to errors whereby they become consonant clusters, while clusters are less prone to errors whereby they become singletons".

tainly does not reflect the conclusions of the studies reviewed above. The majority of studies that found frequency effects in speech errors—including his own when controlling for the Addition Bias—explicitly primed contextual errors (Goldrick, 2002, 2004; Goldrick & Larson, 2008; Levitt & Healy, 1985; Motley & Baars, 1975). With regard to non-contextual errors, in contrast, the evidence is mixed: some studies found frequency effects (Edwards et al., 2004; Tremblay et al., 2016), one an anti-frequency effect (Santiago et al., 2007), and some found no effect of frequency at all (Croot & Rastle, 2004; Munson, 2001; Shattuck-Hufnagel & Klatt, 1979).[9] Likewise, Stemberger's claim that "the anti-frequency effect is not observed with nonce words, though it is consistently observed with real words" (Stemberger, 2004; p. 419) is at odds with the data summarised above, in which several word studies did not find an effect (Munson, 2001; Shattuck-Hufnagel & Klatt, 1979), one found an anti-frequency effect, and his own study found it only in a specific situation (a non-specified phoneme competing with a specified one).

It is thus not entirely clear under what exact circumstances a frequency effect (or an anti-frequency effect) emerges. Previous research suggests that it depends on the method applied and the specific task, in other words, the processing requirements, as well as the dependent variable. Trivial tasks, like simple naming and repetition, seem to produce null results more often than cognitively more demanding tasks, like natural speech production, tongue twister production, syllable manipulation, or tasks involving priming, such as SLIP. The effects found even vary with the time course of processing within a task. While effects of word, syllable, and phoneme frequencies were found in immediate naming tasks, they were absent in delayed naming (Bürki et

[9] Note that only Croot and Rastle (2004) found no effect at all, while Munson (2001) found no effect in adults and Shattuck-Hufnagel and Klatt (1979) no effect of direction in substitution errors but one of general participation in speech errors, as mentioned above.

al., 2015; Mooshammer et al., 2015), which suggests that the effect occurs at earlier processing stages, for instance, lexical access or phonetic encoding. If they were located at the stage of motor programming/execution, there should be no difference between immediate and delayed naming. On the other hand, the presence of effects of sublexical frequencies in nonce words rules out the possibility of frequency effects manifested solely at the stage of lexical access. In contrast to accuracy differences between HF and LF items, which can usually be found, there is often no frequency effect visible in neurological patterns (Brendel et al., 2008; Bürki et al., 2015; Riecker et al., 2008; Tremblay et al., 2016). Brendel et al. (2008) found no neurological effect of syllable frequency, but a behavioural interaction between syllable frequency and complexity: complexity in syllable structure had an effect on reaction time for HF but not LF syllables. Frequency-based differences in latencies emerge in about half the studies, often under some configurations but not others (e.g., Andrews, 1992; Mooshammer et al., 2015). In comparison, articulatory differences between HF and LF items, for example in duration or degree of overlap, are quite reliable (e.g., Munson, 2001; Pouplier, Marin, Hoole, et al., 2017).

With respect to the linguistic units studied, words, phonemes, and transitional probabilities show frequency effects most consistently. Bigrams also display frequency effects more often than not, while syllable effects only occur in a minority of studies—also depending on the methods and dependent variables used. Note, though, that the role that the unit under consideration plays in the language tested may also be decisive for the occurrence of frequency effects. Previous studies have shown that, in general, whether effects of linguistic units emerge during processing depends on their status in the language. For example, syllable priming effects are generally stronger in Mandarin, which has a relatively small number of different syllables (disregarding tone differences) and no resyllabification, than in English, which has a far larger inventory of syllables that are notoriously prone to resyllabification (cf. Cholin et al., 2006). Mandarin speakers, therefore, show no effect of onset preparation, in contrast to English speakers (O'Seaghdha et al., 2010; see also their explanation for these differences in terms of a *proximate units principle*). However, in many studies, a frequency effect in only one unit is investigated and frequencies in the other units are not always controlled for (although see Andrews, 1992; Mooshammer et al., 2015; Santiago et al., 2007; for studies that show effects for multiple units; as well as Cholin & Levelt, 2009; Cholin et al., 2006; for studies that control for various frequency measures), which makes it difficult to compare and assess the different methods used in the various studies. It would be insightful to examine the effects of frequencies of various units relative to each other and compare their influence under different processing requirements.

Finally, some evidence suggests that sensitivity to frequency (or specifically phonotactic probability) depends on the stage of linguistic development. Munson (2001) found an effect of the frequency of heterosyllabic consonant clusters on both repetition accuracy and cluster duration in children, but not in adults. (However, he found more variability among LF clusters than HF clusters in adults as well.) He interprets this as an indication that children's motor representations are more bound to the frequency of the phonological context, while adults' representations are more abstract. A similar study with bigram frequencies not restricted to consonant clusters revealed that vocabulary size was the source of the group difference (children vs. adults) with respect to the effect of bigram frequency (Edwards et al., 2004). Hence lexical development and potentially also the capacity for linguistic abstraction seem to further determine the occurrence of frequency effects.

Summing up, frequency effects have been observed on all linguistic levels, using different paradigms and different dependent variables. They are usually facilitative, meaning HF units are processed more quickly and/or accurately. However, some studies fail to find frequency effects and a few report anti-frequency effects (i.e., inhibitory effects of frequency). Whether a frequency effect is observable depends on many factors, such as the dependent variable (error rates and reaction times typically yield frequency effects, while the results for brain response patterns are more mixed), the task applied (immediate vs. delayed naming), the stimuli used (words vs. nonwords), the language (results for studies on Spanish (L1) speakers contrast with results for other languages), and probably the interaction between these factors, too (e.g., the combination of language and unit studied has proven relevant, which shows that not all linguistic units are equally important or psychologically real in every language). Moreover, it depends on the presence or absence of stronger effects in the opposite direction.

8.2.2. Sonority sequencing and speech production

Most evidence for facilitation of syllables that conform to the SSP comes either from acquisition studies or studies on the productions of individuals with acquired speech impairments. In the productions of healthy adult speakers, effects of sonority sequencing are rarely visible.

During L1 acquisition, children have been observed to reduce consonant clusters to the consonant that results in the least complex syllable in terms of sonority sequencing: the one with the greatest syllable-initial sonority rise and the smallest syllable-final descent (Barlow, 2005; D. K. Ohala, 1999). This means that, in onset clusters, the least sonorous segment of the target cluster is preserved, thereby creating a sharp rise in sonority from that segment to the following vowel.[10] A more complex pattern was found in a study on Greek children's productions (Tzakosta, 2009). They displayed different preferred re-

[10] Barlow (2005) noted one exception to this rule, namely heterosyllabic nasal-voiced stop clusters, which reduced to the more sonorous element in one individual.

pair strategies for different consonant clusters (including affricates), which the author attributes to varying levels of cluster cohesiveness—ultimately related to their sonority profile, making them more or less likely to be perceived as complex segments. According to her, clusters that are phonologically coherent, that is, satisfy sonority sequencing, are perceived as complex segments and undergo limited repair strategies, whereas phonologically less coherent clusters are perceived as true clusters and are repaired frequently via various strategies. She assumes different underlying phonological representations for the various cluster types. Obstruent–liquid clusters with an ideal sonority profile are considered to be the least coherent and show the greatest variety of repair strategies during L1 acquisition. Affricates are regarded as the most coherent "clusters" and are mainly reduced to the stop part. These diverging results could be due to the greater number and complexity of target clusters in Greek than in the languages studied by D. K. Ohala (1999) and Barlow (2005), that is, English and Spanish, respectively. The types of clusters consequently varied across the three studies. In spite of the diverging results concerning reduction patterns, however, it is noteworthy that all of the children studied show sonority effects of some kind. In one study, effects of sonority and phoneme frequency on the order of acquisition were compared directly (Stites et al., 2004). There were no effects of sonority on the order in which onset consonants were acquired. With respect to coda consonants, one of the two children studied showed a preference for frequency over markedness (defined in terms of sonority), while the other proved less sensitive to frequency and acquired the more sonorous coda consonants (i.e., those that create a less marked syllable according to the Sonority Dispersion Principle) first instead.

L2 production seems to be influenced by sonority sequencing in a similar manner. The fine-grained sonority differences between stops and fricatives, as well as voiceless and voiced stops, are represented in varying error rates of L2 learners of English for syllable-initial consonant clusters beginning with these obstruents followed by /r/ (Broselow & Finer, 1991). Crucially, in their study, none of these clusters were permitted by the learners' native languages, Japanese and Korean, but the most well-formed cluster (/pr/) yielded error rates comparable to those for an L1-legal cluster of optimal sonority distance (stop–glide). This suggests that the transition from their restrictive L1 systems to the more lenient target language, English, was guided by sonority evaluations.

Like children and L2 learners with their underdeveloped systems of the target language, patients with acquired language impairments often simplify marked sonority structures in their productions. The most conclusive results come from individuals with apraxia of speech (AOS, e.g., Romani et al., 2011; Romani et al., 2013). There are mixed results concerning the influence of sonority sequencing on aphasic speech production: a number of studies found clear indications of production accuracy in aphasic patients (with impairments as diverse as Broca's, Wernicke's, and conduction aphasia) varying as a function of sonority sequencing (Miozzo & Buchwald, 2013; Romani & Calabrese, 1998; Romani et al., 2013). For example, Romani and Calabrese (1998; p. 98) remark that "[d]eletion rates follow a sonority-based hierarchy of syllable types remarkably well". On the other hand, Romani et al. (2011) and Romani et al. (2013) only found sonority sequencing effects in the productions of patients with articulatory planning deficits, not those with phonological deficits. Beyond the general effect of higher error rates for marked sonority structures, studies on speech production in aphasic and especially apraxic patients specifically found their speech errors to improve syllable sonority profiles (Buchwald, 2009; Miozzo & Buchwald, 2013; Romani & Calabrese, 1998; Romani et al., 2013). However, it was not always the case that productions improved the sonority profiles of the target words. The productions of one patient with a phonological impairment improved sonority sequencing in 44% of cases and aggravated it in 41% of cases (Miozzo & Buchwald, 2013; the remaining errors left sonority profiles unchanged), and Romani et al. (2011) found improvement of sonority profiles only in apraxic speakers but not in speakers with a phonological impairment. Romani et al. (2013) observed an unusual pattern: patients (phonological and apraxic) did not show a tendency for more errors on words with difficult sonority profiles, but when they did produce errors, the sonority profile was improved much more often than not. A number of studies report that the sonority profiles of most productions remain unchanged compared to target words, but the few changes that do occur are in the direction of sonority optimisation (Christman, 1994; Kohn et al., 1998).[11] Hence, although the effect is not completely consistent, studies that compare the syllable structure of phonemic paraphasias with target lexemes showed a tendency towards simplification, not only in general syllable structure but also in terms of sonority profiles. Likewise, there is a bias for simpler sequences when compared to the lexicon of the target language in aphasic neologisms (i.e., productions where no target lexeme can be identified) (Stenneken et al., 2005).

It is clear, of course, that sonority is only one factor that constrains the productions of aphasic and apraxic patients and modulates the processing of sound sequences more generally. Even by its proponents, it is not normally claimed to be the only factor, yet its influence is rarely compared to that of other factors directly. There are a few notable exceptions, however. Most importantly, Romani et al. (2013) observed strong effects of segment frequency, sonority, and markedness in apraxic speech errors and found that, when disentangled from frequency, sonority shows the strongest effects. That is, far more errors improved the sonority profile of a target word while segment frequency decreased than the other way around. The similarly prominent role of sonority in relation to language-specific sequencing preferences became apparent in a large corpus of English and German non-lexical

[11] with the exception of segment additions in Christman (1994), which more often deteriorate sonority contours

aphasic speech automatisms (Code & Ball, 1994). Here, there were no cases in which language-specific phonotactics supersede the SSP. This means that, although aphasic speech generally follows language-specific phonotactics, phonotactic sequences which violate sonority are avoided. In contrast, Stenneken et al. (2005) note that, in spite of the high overall compliance with sonority principles in their patient's (a German Wernicke aphasic) production data, his deviations from the general pattern may be due to a relatively high number of syllable-initial /sp/ and /st/ productions. This deviation is interesting because it reflects the frequency of these German exceptions to sonority principles and thus the potential interaction of sonority and frequency principles driving the patient's productions.[12]

Further examples of other principles overriding sonority principles include (not further specified) constraints that operate on individual segments (Romani et al., 2013) and segment position (Romani et al., 2011). The data in the latter study showed that onset cores (what corresponds to C2 position in the sibilant–stop clusters and C1 position in all other clusters used in the present study) are far more resistant to errors than other positions. All of the above suggests that sonority principles govern phonologically impaired speech to some degree but can nevertheless not be said to be the most important factor in all cases. In a comparison of sonority and language-specific phonotactics, sonority seems to exert the stronger influence (Code & Ball, 1994; Romani et al., 2013), although some results indicate that language-specific frequencies can diminish its effect.

Moreover, the presence of a sonority sequencing effect seems to depend to some degree on the specific impairment of the patient. Differences in sonority sensitivity between groups of patients with impair-

[12] It is startling, though, that /sp/ and /st/, rather than /ʃp/ and /ʃt/, are noted by Stenneken et al. (2005) to be over-represented. It can be speculated whether this shows a bias for a universally unmarked phoneme while keeping the natural class of the L1-frequent marked phoneme intact.

ments on different levels have been recruited as a testament to the level in speech production on which sonority exerts an influence. Based on the locus of the patients' impairments as displaying the strongest sonority effects in their studies and/or diverging behaviour in different tasks, researchers have argued for the phonological lexicon (Kohn et al., 1998; Romani et al., 2011), phonological encoding (Bastiaanse et al., 1994; Stenneken et al., 2005), and articulatory planning (Romani & Calabrese, 1998) as the level at which the effects of sonority sequencing arise. Christman (1992; p. 244) assumes that sonority may be distributed "throughout the entire language system" and accessed during different stages of language production, and Bastiaanse et al. (1994), although isolating the effects to the phonological level, also acknowledge some degree of mutual influence between the phonological and phonetic levels. Moreover, the diverging conclusions of different researchers—for example, Romani and Calabrese (1998) assume the level of articulatory planning, while Stenneken et al. (2005) explicitly exclude this level—suggest that the influence of sonority on speech production is not limited to a single processing level but can manifest itself on several levels. Why is the level relevant to the present study? If the sonority sequencing effect were located (solely) at the level of the phonological lexicon, it would not show up in this experiment, which uses only pseudowords and no existing lexical items. A locus on the level of phonological encoding, articulatory planning, or a distribution over several levels, on the other hand, make the occurrence of a sonority sequencing effect in the data collected for this experiment more likely.

In the same vein, the mechanism by which sonority sequencing is implemented is of interest. Different accounts have been presented. According to J. J. Ohala (1992), sonority sequencing is merely an artefact of speech production. Lindblom (1983) corroborated this stance with an analysis of Swedish consonants in terms of jaw movements showing a strong correspondence with sonority. Some researchers claimed that the effects of sonority sequencing are so strong and incorruptible because sonority is hard-wired into the brain, although the postulated implementations (e.g., Sussman, 1984; p. 169: "each consonant and vowel position is associated with a specific cell assembly network") seem highly implausible or even untenable from a modern neurolinguistic point of view. If sonority sequencing has a mainly articulatory basis, it is likely to affect speeded production—as used in the present experiment—such that preferred syllables are kept intact, while dispreferred syllables are optimised sonority-wise. If, on the other hand, sonority has neuronal correlates, it might only show effects in certain tasks.

In contrast to aphasic/apraxic productions, data from healthy individuals hardly ever show any influence of sonority sequencing. In a study by Stemberger (1991), neither spontaneous nor elicited slips of the tongue display sonority effects and he concludes: "The sonority hierarchy fails to predict some asymmetries and wrongly predicts others. Sonority is not relevant here [...]". In a metalinguistic task, initial consonant clusters were treated the same by subjects irrespective of their sonority profiles (Treiman, 1986; there was a sonority effect on syllable-final clusters, though, which might be taken as an indication of the greater inherent cohesiveness of onsets). Furthermore, in contrast to cluster frequency, a cluster's sonority profile does not affect a its disposition for articulatory overlap under rate changes (Pouplier, Marin, Hoole, et al., 2017).

Summing up, there is ample of evidence for an influence of sonority sequencing on impaired speech production, as well as speech production during language acquisition. In most of the studies discussed here, the sonority *distance* between adjacent phonemes was optimised rather than merely violations of the SSP "repaired". In contrast, sonority sequencing seems to have very little effect on fully-developed, healthy speech production. There are two possible explanations for this discrepancy: either sonority only exerts an influence on developing or damaged—i.e., somehow weak—language production systems, or it can potentially also affect healthy mature systems; but ceiling effects prevent visible effects in most cases, so the task must be more difficult for them to emerge.

Effects of sonority sequencing are rarely directly compared to those of language-specific distributions, but in the rare cases that do exist, effects of sonority seem to outweigh those of language-specific phonotactics in clinical studies (Code & Ball, 1994; Romani et al., 2013), while frequency effects prevail in studies with non-impaired speakers (Stemberger, 1991). Evidence from language acquisition studies is mixed, with some children guided by frequency and some by sonority (Stites et al., 2004).

8.2.3. Phonological neighbourhoods in speech production

Another factor that has proven to be important in the speech production process are lexical neighbourhoods.[13] Studies that investigate the effects of neighbourhood density find that the number of phonological neighbours a word has affects its production accuracy (Vitevitch, 2002), production ease (as evidenced by tip-of-the-tongue states; Harley & Bown, 1998; Vitevitch, 2003), or latency (Andrews, 1992; Buz & Jaeger, 2016). Only under specific circumstances do neighbourhood effects fail to emerge. For example, Andrews (1992) found them in a standard naming task but not in delayed naming, and Sadat et al. (2014) observed an effect on reaction times but none in the error analysis.

Although neighbourhood effects are relatively reliable in studies of speech production, there is some uncertainty about their direction, especially when it comes to sublexical units. While neighbourhood ef-

[13] Since the task in the present experiment is purely oral and does not involve reading, only findings concerning phonological neighbourhoods (as opposed to orthographic neighbourhoods) will be presented here.

fects are overwhelmingly inhibitory in speech perception because different lexical items corresponding to the auditory input compete for recognition, speech production research has unearthed both facilitative effects of lexical neighbourhoods (Andrews, 1992; Harley & Bown, 1998; Vitevitch, 2002, 2003), which suggests a supportive role of similar lexemes in the production process, and inhibitory effects (Sadat et al., 2014; Vitevitch & Stamer, 2006), which indicates competition between them. Buz and Jaeger (2016) found a facilitative effect on reaction times and an inhibitory effect on word durations; this means, words with large neighbourhoods were named faster but were longer in duration; and Vitevitch and Luce (2016; p. 7.12) conclude that "it is not clear what factor (or factors) determines whether phonological neighbors facilitate the retrieval of or compete with a phonological word form during speech production". However, Sadat et al. (2014), who found inhibitory effects in their own experiment, reanalysed a number of previous studies using finer statistical methods and found converging evidence for an inhibitory effect on response latencies. They note that this only applies to unimpaired production, though, while facilitative effects can emerge in cases of disrupted speech production. According to them, the "phonological neighborhood generates two opposite forces, one facilitatory and one inhibitory", and "inhibitory processes dominate in efficient naming by healthy speakers" (Sadat et al., 2014; p. 33). A distinction that might have contributed to the diverging results is made by Stemberger (2004). He categorises neighbours into *friends* and *enemies*. Friends are defined as neighbours that share a certain characteristic of the target word, for example, a feature primed in a SLIPs experiment. Enemies, on the other hand, are neighbours which do not share the target characteristic in question. According to Stemberger (2004), only friends have a (facilitative) effect on word production, while enemies or the overall neighbourhood size do not. On the whole, the metrics of lexical neighbourhood density and summed neighbourhood frequency might, therefore, be too much of a simplification to accurately capture the processes and phenomena at work during the production of a word (see also Vitevitch & Castro, 2015; for a detailed account of why and how further characteristics of the organisation of the mental lexicon and a word's position in it have to be taken into consideration).

Another important aspect is that neighbourhood effects are thought to operate on the lexical level. Vitevitch and Luce (2005; p. 194, citing Vitevitch and Luce, 1998) summarise that "only stimuli—either words or nonwords—that activate and resonate with lexical representations will produce a neighborhood density effect". In contrast, the experiment described here investigates sublexical processing. The standard measure of lexical neighbours is therefore inadequate for the present purposes, and an equivalent on the sublexical level is more useful. For this reason, the neighbourhood measure was adapted to the sublexical level in the experiments reported in this dissertation. Nevertheless, it is not clear whether the competitive effect of lexical neighbourhoods applies equally to the sublexical level. The notion of sublexical neighbours is also taken up in Cholin et al.'s (2011) exposition of the mental syllabary, where it is assumed that not only the motor programme for one syllable but a whole "syllable neighbourhood" becomes activated during production. However, they remain agnostic as to whether syllable neighbourhood has a facilitative or an inhibitory effect and do not examine its role directly. The direction of a potential effect of neighbourhoods on the subsyllabic level is equally unclear.

To sum up, neighbourhood effects in speech production are complex and depend on various factors. Moreover, they have—to the best of my knowledge—been studied exclusively on the lexical level, that means, as whole lexemes, which might potentially receive activation from the lemma level. It is questionable whether neighbourhood effects at this level will be induced at all by the kind of stimuli (pseudowords) and the task used in the present experiment. Oral repetition can be performed without lexical processing, which means without activating potential lexical neighbours. On the sublexical level, which is the focus of the present study, neighbours might become activated, but sublexical neighbourhood effects could be different from lexical ones. There is, as yet, not enough research on their role in the speech production process to make well-founded predictions concerning their effect; but the present experiment will hopefully shed some light on that matter.

However, neighbourhood effects seem to be weaker than effects of phonotactic probability. When Vitevitch and colleagues (Vitevitch et al., 2004) manipulated phonotactic probability and neighbourhood density of words separately, only phonotactic probability showed a significant effect in a picture naming experiment. There was no main effect of neighbourhood density and no interaction between the two, which suggests that the sublexical level dominates in speech production.

8.2.4. Ease of Articulation

Obviously, articulatory factors also influence speech production accuracy and present potential confounds to the main variables in this study. Articulatory complexity has proven a significant factor in word production both in healthy populations and individuals with speech disorders. For example, increased articulatory complexity leads to a decrease in word repetition accuracy in patients with AOS (Bislick & Hula, 2019; Romani & Galluzzi, 2005; Ziegler & Aichert, 2015), to a decrease in production accuracy under high cognitive load in healthy individuals (Pouplier et al., 2014), and to increased neurophysiological recruitment of the basic speech network in healthy individuals (Bohland & Guenther, 2006). In contrast, it failed to show effects on naming latency and accuracy in healthy individuals (Levelt & Wheeldon, 1994). What further complicates interpreting the effects of articulatory complexity is that the concept is not uniformly defined. Depending on what a researcher defines as “complex”, the results might vary considerably. A well-defined formalisation is Ease of Articulation (EoA, Ziegler & Aichert, 2015), which attempts to capture “the phonetic planning costs for the production of spoken words from their gestural architecture" (Ziegler & Aichert, 2015; p. 27). EoA is largely based on Articulatory Phonology (and, as such, takes phonetic gestures as the core units of speech production) and comprises articulatory characteristics of a given target, such as word length, metrical structure, presence or absence of consonant clusters, as well as glottal or velar aperture gestures and lip or tongue gestures. EoA has been shown to reliably predict word and nonword production accuracy in aphasic speech production (Ziegler & Aichert, 2015).

8.2.5. Other factors in speech production

In addition to language-specific phonotactics, sonority, and articulatory aspects, other factors can influence speech production more generally and the occurrence of speech errors in particular. Here, only the two factors most relevant to the experiment at hand will be briefly discussed.

The first concerns the featural similarity of phonemes involved in an error. In substitution errors, there is usually a high degree of similarity (as measured by the number of shared features) between the target phoneme and the substituted phoneme (Levitt & Healy, 1985; Shattuck-Hufnagel & Klatt, 1979; Wilshire, 1998, 1999). For contextual errors, this means that, the more similar the phonemes in a target sequence are, the more likely errors will occur (see also Section 7.3). This is intuitively plausible and also easily explicable within connectionist frameworks of speech production, in which phonemes receive activation from the featural level. A higher error rate can therefore be expected for stimuli that contain similar phonemes in (e.g., initial /ts/ and /ks/ or /fl/ and /sl/).

Another structural effect that has been found in speech error analysis is one of consonant position—C1 vs. C2—in an initial consonant cluster (Stemberger & Treiman, 1986).[14] This effect was found when a cluster interacted with another cluster, but not when it interacted with a singleton consonant. Both in a corpus of naturalistic speech errors and in a SLIP experiment, there was a tendency in deletion errors to delete the second consonant in an initial cluster rather than the first[15] and a tendency in addition errors to insert consonants into this position. The authors ascribe this to a "difference in accessibility of the first vs the second consonant of a cluster" (Stemberger & Treiman, 1986; p. 176) and conclude that the singleton consonant of a simple onset takes the same position as C1 in a cluster, a position that possesses greater inherent activation than C2. According to the authors, the fact that C1 position has a higher activation than C2 might have something to do with the fact that it—as a structural position—has a higher frequency than C2. For the current experiment, this means that a bias for C2 deletion, irrespective of the relative frequencies of the two consonants in relation to the cluster as a whole or the sonority profile, should be expected. In cases of addition errors, consonants are unlikely to be inserted into C1 position.

8.3. Research questions and hypotheses

The central research question of this study is whether language-specific distributions or universal phonological principles primarily determine which initial consonant clusters are the most difficult and which ones are the easiest to produce for adult native speakers. If the usage-based assumption that our use of language shapes mental representations and influences later processing is correct, then

[14] The effect was slightly smaller when C1 was /s/ as opposed to a voiceless stop, though.

[15] In many instances, this tendency can be explained in terms of sonority optimisation (see Section 8.2.2), but the fact that the effect was also found with s-initial clusters indicates that the skeletal position is the main cause.

language-specific distributions, operationalised here as cluster frequencies, should show the strongest effect. If, on the other hand, universal principles of phonological well-formedness are more relevant to production effort, then a clearer effect of sonority should be visible, which is known to influence not only the lexicons of languages but also L1 acquisition and (at least) impaired speech production. In this experiment, the relative influences of these two forces on speech production are tested using a tongue twister paradigm. Particular attention will be paid to consonant clusters for which the predictions made by the two principles diverge. Error rates will serve as the operationalisation of production difficulty.

Based on previous research and usage-based theories of language, it is hypothesised that frequencies on the level of consonant clusters will influence production accuracy; this has been found, for example, on the phoneme or syllable level or for phonotactic probability in general. More specifically, the hypothesis is that more errors will occur on LF clusters than on HF clusters and that errors will tend to create HF clusters.

With respect to sonority, predictions are more difficult. Most of the observed sonority sequencing effects come from studies on L1 acquisition or impaired speech production. Evidence for a sonority sequencing effect in unimpaired speech production is sparse, but it is possible that task demands in previous studies have not been sufficient to evoke such a response in healthy speakers. Therefore, a very demanding task is used in the present experiment in order to investigate whether sonority sequencing can also shape unimpaired speech production under certain circumstances. Moreover, it has been shown that, in fast speech, syllabification conforms more to universal sonority constraints, while in slow speech, language-specific syllabification rules can override sonority-based principles of syllabification (Laeufer, 1995). As the speech rate in the present experiment is relatively high, it is possible that sonority effects become more visible here than they are at slower speaking rates.

It is therefore hypothesised that consonant clusters with a small sonority distance, especially those that violate the SSP, will be more vulnerable to speech errors than clusters with a large sonority distance. Furthermore, errors are likely to improve the sonority profile of a consonant cluster in terms of the SSP and the SDP.

However, cluster frequency is hypothesised to be more influential than sonority. This means, particularly with respect to the clusters for which predictions diverge, error rates and error outcomes are probably better predicted by frequency than by sonority.

Since speech processing involves dynamics of activation and competition in a complex network of phonological representations, it is plausible that not only the frequencies of the target clusters influence their production but also the characteristics of their phonological neighbourhoods. Previous studies have yielded conflicting results concerning the direction of the neighbourhood effect in speech production, so it is not possible to predict a neighbourhood effect or its direction with certainty. If neighbourhood effects are transferable from the lexical to the sublexical level, however, it is more likely that the effect of neighbourhood frequency will be inhibitory (i.e., clusters with many and frequent neighbour clusters will show higher error rates) because this study is concerned with unimpaired speech production or that there will be no effect of neighbourhoods that are undifferentiated in terms of friends and enemies.

8.4. Methods

8.4.1. Participants

Forty-one young adults (26 female; mean age: 22.92; SD = 3.48) participated in the study and received monetary compensation for their participation. All of them were native speakers of German and reported speaking no German dialect. Again, participants who grew up in the south of Germany were excluded from the study in order to avoid conflicting frequency information with respect to the cluster /ks/ due to dialectal vowel deletion, for example, in *g'sagt* /ksaːkt/ "said". Six of the subjects reported having undergone speech therapy in the past and one subject reported a speech development disorder that was not treated but disappeared over time. All subjects gave written informed consent to participate in the study.

8.4.2. Materials

Eighty pairs of monosyllabic pseudowords with CCVV, CCVː or CCVC structure served as stimuli. Pseudowords rather than real words were chosen to eliminate lexical influences and isolate the effects of the sublexical variables of interest. Moreover, pseudowords have been shown to elicit a higher error rate than real words (Wilshire, 1998). It should be kept in mind, however, that this decision has implications for the presence vs. absence of certain effects. The same 16 consonant clusters as in the perception experiments were used as onsets for the stimuli. The consonant clusters were arranged into ten pairs (see Table 8.1) of minimally different clusters: clusters that either differ in one feature in one of the consonants, (e.g., /tr/ and /kr/ differing in place of articulation in C1) or that are composed of the same two consonants in reversed order (here called *metathesis pairs*, e.g., /sk/ and /ks/). As can be seen from Table 8.1, four consonant clusters (/ʃt/, /ks/, /ps/, /fl/) appeared in two pairs. This was done to maximise the competitor contrasts (and thus also the possible error outcomes), as well as the possibilities for individual comparisons. These pairs of consonant clusters served as the basis for the stimulus pair of a trial. For each cluster pair, eight stimuli were created.

cluster pair	frequency difference	sonority difference	stimulus example
ʃt–tʃ	0.94	yes	ʃtœf tʃaf
ʃt–ʃn	0.65	yes	ʃtoː ʃnyː
ʃp–ʃm	0.18	yes	ʃpɛl ʃmɛl
ks–sk	0.24	yes	ksɔʏ skaʊ
ps–sp	0.15	yes	psuːk spuːx
fl–sl	1.05	no	fliːm sloːn
pl–ps	0.97	no*	plɪf psɪç
ts–ks	0.71	no	tsɛm ksaɪn
tr–kr	0.17	no	trɛl krɛŋ
ʃl–fl	0.00	no	ʃleːç fløːç

Table 8.1.: Pairs of consonant clusters used in stimuli (HF clusters first)
Frequency difference shows the difference in log type frequencies between the first and the second cluster
*no difference in SSP violation; difference in sonority distance: 2

All stimulus syllables conformed to German phonotactics and included all German vowels and diphthongs (/aː, a, eː, ɛ, iː, ɪ, oː, ɔ, uː, ʊ, øː, œ, yː, ʏ, aɪ, ɔʏ, aʊ/), as well as all licit simple codas (/p, t, k, f, s, ʃ, ç, x, m, n, ŋ, l, ɐ/). Since two phonemes are more likely to be involved in an error if they appear in the same environment (Dell, 1988; p. 134), stimulus pairs with identical vs. different vowels were balanced across onset clusters. In all stimulus pairs except those with onset /ts/–/ks/, half of the stimulus syllables had an identical vowel, while the other half differed in vowel. For /ts/–/ks/ stimuli, 10 stimulus pairs had different and 6 identical vowels. For each cluster pair, in approximately one third of the stimuli (i.e., either five or six stimuli), the two syllables shared the whole rime, whereas in the other two thirds, the rime differed between the two syllables. Seven syllables (/flɛm/, /kseːl/, /ksɛl/, /psɪç/, /ʃluː/, /ʃtɪŋ/, /tʃaː/) occurred in more than one stimulus pair. In addition to the 80 test item pairs, 50 pairs of filler items with CVː, CVV, VC, or CVC structure were constructed. Two lists of stimuli were created. The order of the two syllables in a stimulus pair was counter-balanced across lists (e.g., /ʃtœf tʃaf/ and /sloːn fliːm/ in list A; /tʃaf ʃtœf/ and /fliːm sloːn/ in list B).

The stimuli were spoken by a female native speaker of German and recorded with an AKG C2000B microphone in a sound-proof booth using Adobe Audition. The recording was saved directly on a computer with a sampling rate of 44.1 kHz (16-bit resolution). All syllables were recorded several times and the best token of each syllable was selected for inclusion in the set of audio stimuli. Using Praat (Boersma & Weenink, 2018), all stimulus syllables (i.e., test and filler items) were then normalised to 65 dB SPL and concatenated by twos with a 500 ms silence between them and a 500 ms silence after the second syllable to form the stimulus pairs.

In order to avoid sequence effects, the order of the stimuli was pseudo-randomised for each participant separately using the software Mix (van Casteren & Davis, 2006). The constraints for pseudo-randomisation were as follows: 1) No more than three test items occurred in direct succession before a filler item intervened. 2) There was a minimal distance of three trials before the same consonant cluster pair was repeated; this was used to prevent practice effects for any particular cluster. 3) The same two cluster pairs could only alternate four times before a stimulus with a different cluster pair occurred. (In reality, no two cluster pairs alternated that often.) 4) Items of the highest articulation difficulty class[16] were separated by at least one trial; this was done to reduce fatigue effects. 5) The same vowel in either syllable 1 or syllable 2 could occur in no more than two consecutive trials.

[16] All stimulus items were rated by the experimenter according to their subjective articulation difficulty on a scale from 1 to 3.

8.4.3. Design and procedure

Prior to the experiment proper, participants completed a questionnaire (see B in the Appendix) and carried out a short speech production task and a digit recall task. The questionnaire contained general questions about the participant, such as age, gender, and field of study, questions on language background (including dialect and foreign language experience, speech and speech development disorders, and therapy, as well as hearing impairment) and on musical training (to control for experience with rhythmical training).

The speech production task comprised of casual reproduction of four sentences printed on a sheet of paper and was conducted to check for regional influences on pronunciation. Participants were instructed to read the sentences silently and then speak them out loud as if they were saying them in an informal conversation. All of the productions were rated as standard-like by the experimenter, so all participants were included in the experiment.

In the digit recall task, participants listened to four rows of six digits each. After each row, they had to repeat all digits of a row orally once. This task was included to control for memory deficits. Performance varied among subjects from 0 to 4 mistakes (mean: 0.8, median: 0). There was no statistically significant correlation between participants' performance in the digit recall task and the experiment task (Spearman's ρ (39) = 0.05, p = 0.77); it can thus be assumed that their errors in the experiment task were not substantially influenced by short-term memory deficits.

During the main experimental task, a tongue twister paradigm was used to elicit speech errors. The tongue twister task was similar to one that has been proven to effectively elicit contextual speech errors (Dell et al., 1997; Vousden & Maylor, 2006), but instead of real words forming a phrase, the pseudoword pairs described above and exemplified in Table 8.1 were used. Stimuli were presented auditorily for repetition because this was considered to be closer to normal speech production than reading them out loud from a screen or paper. In so doing, artefacts from reading were excluded. During each trial, participants heard a stimulus pair twice over headphones at a slow pace while the screen remained black. Shortly after the offset of the last stimulus syllable, a white fixation dot appeared in the middle of the screen to indicate that the participant should prepare to speak. The following production phase was divided into a *familiarisation phase* and an *elicitation phase*. During the familiarisation phase, participants had to repeat the stimulus sequence once at a pace of 63 beats per minute (bpm). The purpose of the familiarisation phase was two-fold: on the one hand, participants were given the chance to produce a stimulus slowly before the challenging task of repeating it quickly (hence "familiarisation"). On the other hand, the recordings of the productions also served to check whether or not the stimuli had been correctly perceived. After this slow production, another white fixation dot appeared in the middle of the screen to prepare participants for the *error elicitation phase*, in which they had to repeat the sequence four times without a pause at a pace of 144 bpm. The speed for stimulus production was indicated by auditory metronome clicks presented to the subjects over headphones. The metronome clicks were recorded from https://www.metronomeonline.com/ with Praat (Boersma & Weenink, 2018) and played back in Open Sesame (Mathôt et al., 2012) during the experiment at original speed. Note, however, that it was not possible to control for whether subjects strictly kept to the predefined speed. It is thus possible that some stimuli were produced slightly more slowly than others. Productions that were noticeably slower than the predefined speed were excluded from analysis. This procedure of one slow and several fast repetitions in time to a metronome has proven fruitful in earlier studies (e.g., Goldrick & Larson, 2008). The experiment was self-paced; after each trial, subjects had to click the mouse in order to start the next one.

Subjects' productions were recorded over an AKG HSD 171 headset that was connected to a Focusrite iTrack Solo. This audio interface was connected to a MacBook Pro on which the audio was recorded in wav-file format in Praat (Boersma & Weenink, 2018). Productions were additionally recorded on a tape recorder with an internal microphone placed about 50 cm from the participant in case something went wrong with the primary recording. This backup was used for annotation of the data of the first two subjects due to technical problems with the primary recording. All the other annotations were based on the primary recording. Subjects were instructed to comment on auditory or memory-related uncertainty (i.e., if they did not perceive the audio stimulus accurately or got confused during the elicitation phase and forgot what the target was) after their productions. Productions followed by such a comment were excluded from the analysis, as were productions from trials in which the productions in the familiarisation phase deviated from the target. Four practice trials comprised of stimuli with a simple CV structure were given to familiarise participants with the task and the rhythm. Participants who were unable to complete the task after the four practice trials were allowed to take them again. This was the case for five participants. The total duration of the experiment varied between participants from around 45 to 60 minutes.

Pretests

The settings described above were determined on the basis of pretesting with seven subjects. None of these subjects participated in the final experiment. The parameters varied in the pretest were:

- the order in which the syllables of a stimulus should be produced (ABAB vs. ABBA) and whether stimuli should contain syllables beginning with singleton consonants as well (e.g., /ʃnɪɐ ʃtuːx ʃɪɐ ʃuːx/ for the onset pair /ʃn/–/ʃt/)

- stimulus presentation mode: auditory only, auditory and visual (as words displayed on a computer screen) simultaneously, or auditory first and visual during the production phase (familiarisation and elicitation)
- repetition rate for familiarisation phase and elicitation phase, with rates tested ranging from 63 bpm to 76 bpm for the familiarisation phase and 138 bpm to 168 bpm for the elicitation phase
- the number of practice trials needed to feel comfortable with the task

8.4.4. Data preparation and analysis

Data preparation

The data resulting from the experiment consisted of 26,240 observations (8 productions per trial × 80 test trials × 41 subjects). Phonological transcriptions of the productions were made according to SAMPA conventions. Since the productions were continuous repetitions of the stimuli for each trial, it was not always clear-cut whether a consonant phoneme belonged to the coda of one syllable or the onset of the following syllable. It was assumed that phonemes were assigned to the correct syllable positions during production, so when there was a coda [f] in one of the target syllables, for example, an [f] production between the nucleus of one syllable and target C2 (or even C1) of the following syllable was counted as belonging to the coda of the first syllable. If one of the target syllables started with an onset [s], on the other hand, an [s] production between the nucleus of the first and target C1/C2 of the second syllable was counted as belonging to the onset. If a produced phoneme was not part of the target sequence in either position, the response was excluded from the analysis.[17] In spite of this information-

[17] Examples: In [flaʊnfʃloːm] for target /flaʊn ʃloːm/, the interconsonantal [f] was counted as belonging to the onset of the second syllable because the stimulus con-

driven and yet relatively conservative procedure, mistakes in assigning produced segments to one of the syllables in a continuous utterance cannot be completely excluded. Some productions were transcribed as *NA*. The following cases led to annotation as *NA*:

1) There was no production for a given target syllable (265 cases).

2) The production was unintelligible (71 cases).

3) The production clearly deviated from the speed indicated by the metronome (82 cases).

4) A given production could not be unambiguously ascribed to one of the two target syllables (19 cases).

5) The onset was already produced incorrectly in the familiarisation phase (1,587 cases).

6) The subject made a comment about perception or memory problems (1,352 cases).

Data from two subjects were excluded from the analysis entirely because they failed to correctly produce more than half of the test items during the familiarisation phase and were clearly not concentrating on the experimental task. Furthermore, two stimuli (/tsaː ksaː/ and /fløː sluː/) in list B were excluded because the wrong audio file was attached for one of the syllables. (The corresponding stimuli /ksaː tsaː/ and /sluː fløː/ in list A were included.) This left 24,637 syllables in the remaining data set. A random sample of approximately 13% of all produced test syllables (3,423 syllables stemming from nine different subjects) was transcribed by a second rater, a trained linguist with experience

tained an /f/ in onset position but none in coda position. In [spaʊfspaʊ] for target /spaʊ psaʊ/, the second syllable was excluded because it is not clear whether its onset is produced as [sp] or [fsp].

in phonological annotations, who was naïve as to the object of the experiment. Inter-rater reliability between onset transcriptions was very high (Krippendorff's $\alpha = 0.932$), so that it can be assumed that the remaining data, transcribed only by the first rater, are reliable and serve as an adequate basis for the following statistical analyses.

After annotation, the onset of each produced syllable was split off and compared to the target onset. The rest of the syllable was not analysed. If target and produced (i.e., transcribed) onset deviated phonemically, this was counted as an error. Otherwise, the production was counted as correct.[18]

In addition to the phonemic transcriptions, phonetic anomalies and hesitations were annotated in a dichotomous way but not described or categorised any further. Phonetic anomalies were observed in 807 productions and hesitations before 517 productions.

Analyses

Two sets of analyses were performed: Firstly, mixed effects logistic regression models with error as the binary dependent variable were run. Since an error is likely to lead to further errors on the same item in the following repetitions (Humphreys et al., 2010), another model was run, in which only the first error of a trial on each syllable was counted.

Secondly, to test the hypotheses on error outcomes, another regression model to predict the substitution of the partner cluster (i.e., the cluster it was paired with in a stimulus; cf. Table 8.1 on p. 265) based on the frequency difference and the difference in sonority between the two partners was fitted. This was done to investigate the role of competition in contextual errors more closely. The model used fre-

[18] It was decided not to categorise production errors into classes like anticipation, perseveration, etc. because, due to the trial design with four repetitions of the same two pseudowords, it would be impossible to say for the majority of cases if an error was caused by an element preceding it or following it.

quency difference between the partner clusters (calculated as log partner cluster frequency minus log target cluster frequency) and sonority improvement (calculated as sonority distance of the partner cluster minus sonority distance of the target cluster) as numeric fixed effects and had random slopes for both fixed effects by subject and a random intercept for stimulus syllable nested under target cluster.

The logistic mixed effects regression models with random intercepts for subjects and items and random by-subject slopes for the frequency and sonority effects were fitted with the lme4 package (Bates et al., 2015) in R (R Core Team, 2016). Model fitting was done by forced entry of the variables thought to be influential based on previous studies and theoretical considerations, as well as one variable whose influence became apparent during the transcription of the experiment data. The variables used for model fitting are log cluster frequency, sonority distance and SSP violation, summed frequency of neighbouring consonant clusters, coda in the previous syllable (making the onset more complex in continuous speech), syllable type frequency, identical coda in both stimulus syllables, and metathesis of consonants in the partner onsets. Regarding the potential effect of sonority sequencing, both SSP violation as a binary variable and sonority distance between the two consonants of a cluster as a numeric variable are plausible predictors from a theoretical perspective. Since the studies reported in Section 8.2.2 mostly investigate sonority distance, this predictor was used in one regression model. The binary predictor SSP violation, as utilised in the perception experiments, was used in another model (otherwise identical to the first model) for better comparability within this dissertation and in order to compare the effects of the two variables with one another.

The variable that was added based on inspection of the raw production data is consonant metathesis within a stimulus pair. Transcription and initial data processing revealed that the cluster pairs composed of the same consonants in reversed order (i.e., metathesis pairs, /ʃt–tʃ/,

/sk–ks/, and /sp–ps/) had by far the highest error rates. To control for this influencing factor and prevent it from skewing the regression lines for the other predictors, it was added as a binary variable (metathesis vs. no metathesis within a stimulus pair). Furthermore, it was entered into an interaction term with cluster frequency because it was hypothesised that its influence would be stronger on LF clusters than on HF clusters.

Single phoneme frequency (i.e., the sum or average of the frequencies of the two consonants in a cluster), although of theoretical interest, was too highly correlated with cluster frequency (Pearson's r = 0.63, p < .001) to be included in the model.[19] After fitting this model, non-significant variables were taken out step-wise, and AICs for the models with and without the predictor were compared until no insignificant predictors were left in the model. The same method was applied to the model for the subset that excluded repetition errors.

All continuous variables were centred and factor variables were sum-coded.

To assess the potential improvement of clusters (in terms of frequency and sonority) in speech errors that result from direct competition within the stimulus pair, log frequencies and sonority distances

[19] When considering the influence that ease of articulation has shown on speech production in other studies, it would have been desirable to include it as a predictor in the present model as well. The most accurate operationalisation that can be found in the literature is that of Ziegler and Aichert (2015), which was developed by weighting the relevant parameters (consonant cluster, velar constriction, tongue tip gesture, etc.) through computational modelling of production data from individuals suffering from apraxia of speech. When applied to the 16 test clusters used in this study, however, only five unique values were represented, with eight very dissimilar clusters sharing one of them. This distribution would hardly yield interpretable results. It is also problematic that the order of consonants does not play a role in this measure: /sk/ and /ks/ yield the same score of 0.28. This does not model the articulatory reality well since "the articulatory movements in /sk/ and /ks/ are not mirror images of each other" (Brunner et al., 2014; pp. 411–412). It was therefore decided not to include EoA in the model. It should be kept in mind that this shortcoming reduces the explanatory power of the model.

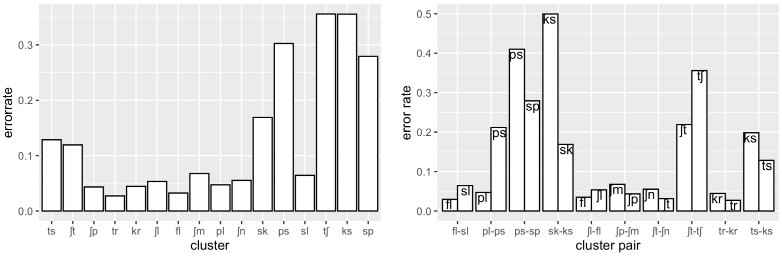
of target and produced clusters were compared, in addition to the main analyses. Only syllables produced with an onset cluster were included in this analysis since the frequency of a cluster cannot be compared to that of a single consonant and simple onsets naturally do not have a sonority distance in themselves. (Recall that, for the target clusters, it was the distance between the two consonants that was measured; the transition to the following vowel was not taken into account.) However, all productions of legal German onsets were considered, not only the set of test clusters.

8.5. Results

Of the 24,637 test syllables in the experiment corpus, 18,750 were produced correctly, 3008 included a production error, and 2879 were annotated as *NA* (for one of the reasons specified above). Leaving aside the NAs, an overall error rate of 13.8% was yielded. Error rates varied between subjects from 1.9% to 29.4%.

Error rates also varied considerably over onset clusters, as can be seen in Figure 8.1a and range from 2.7% for /tr/ to 35.6% for /ks/ and /tʃ/. However, they also diverged greatly within the same target cluster depending on which cluster pair it appeared in (see Figure 8.1b). The three clusters that are part of two different pairs had a much higher error rate in the metathesis pair (i.e., in a stimulus together with their reversed counterpart) than in the non-metathesis pair. For example, /ks/ had an error rate of approximately 50% when paired in a stimulus with /sk/ but only approximately 20% when paired with /ts/.

Errors were significantly less frequent for the first production of each trial when compared to the following three productions, χ^2 (3, N = 21,758) = 539.85, $p < 0.001$.



(a) Error rates over individual target clusters

(b) Error rates over clusters grouped by pair

Figure 8.1.: Error rates over target clusters

Around 42% of the error productions were phonotactically illegal (non-native singleton onset /s/[20], /θ/, and /ɬ/ as well as illegal onset clusters), which supports recent findings that speech errors do not obey phonotactics as much as previously thought. Moreover, 3.7% of all produced onsets (including target and non-target phonemes) were phonetically anomalous. Phonetic anomaly was determined by auditory inspection and included phonemes that were considerably shorter in duration than normal productions (sometimes hardly audible), segments that contained characteristics of two different phonemes (e.g., intensity peaks at several frequency bands, indicative of simultaneous constriction gestures at two different places of articulation), and stops whose VOTs were in between those for voiced and voiceless stops, among others.

About half of the error onsets were two-consonant clusters, one sixth singleton consonants, and the remaining third clusters consisting of more than two consonants. This means that addition errors were more than twice as frequent as deletion errors (cf. also Table 8.2 for the distribution of error types). Many of the produced onsets that consisted of more than two consonants contained realisations of both competing

[20] Since dialect speakers were excluded from the experiment and /s/ is not permitted in syllable-initial position in Standard German, it was counted as illegal.

error type	proportion
addition	33.4%
deletion	14.2%
substitution	
internal	39.0%
external	13.4%

Table 8.2.: Distribution of error types (*Internal substitution* denotes substitutions with the partner cluster.)

consonants (e.g., /fsl–/ in target /fl–sl/ pairs or /sps–/ in /sp–ps/ pairs), which contributed to the high number of illegal onsets.

Figure 8.2 displays how often each of the test clusters was produced as the result of an error.

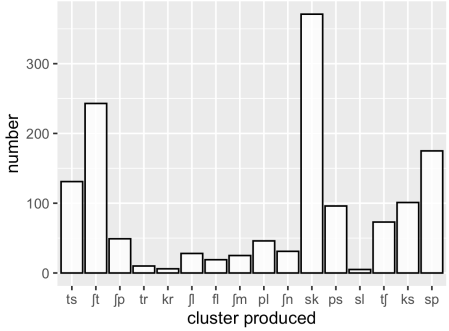


Figure 8.2.: False positive rates of the test clusters

8.5.1. Logistic regression

The final model for the complete dataset included log cluster frequency, sonority distance, metathesis, complex cluster (i.e., a coda in the previous syllable), and identical coda in both syllables and an interaction between cluster frequency and metathesis as fixed effects. The model summary is displayed in Table 8.3. The effects of (summed) cluster neighbourhood frequency and syllable type frequency were not significant in the bigger models; these terms were therefore removed. Identical models were run for the other three cluster frequency measures: CELEX-based token frequencies, type frequencies extracted from the Elexiko dictionary (https://www.owid.de/docs/elex/start.jsp), and token frequencies based on television subtitles (Marian et al., 2012; CLEARPOND). While the type frequencies from Elexiko yielded similar results as those from CELEX, none of the token frequency measures had a significant effect on error rates. Furthermore, it should be noted that all effects in the GLMMs, except those of metathesis and coda in previous syllable, were very unstable and varied considerably as other predictors were added or removed.

predictor	estimate	SE	z value	
(Intercept)	-2.64800	0.17746	-14.922	***
log cluster frequency (type)	-0.29352	0.12158	-2.414	*
consonant metathesis	-0.99069	0.09479	-10.451	***
sonority distance −1	-0.13862	0.20960	-0.661	
sonority distance 1	0.72749	0.18909	3.847	***
sonority distance 2	-0.46886	0.23997	-1.954	.
coda in prev.	-1.28444	0.07366	-17.437	***
coda identical no	-0.56257	0.07508	-7.493	***
coda identical yes	-0.62626	0.08299	-7.546	***
log cluster freq × metathesis	-0.20198	0.05758	-3.508	***

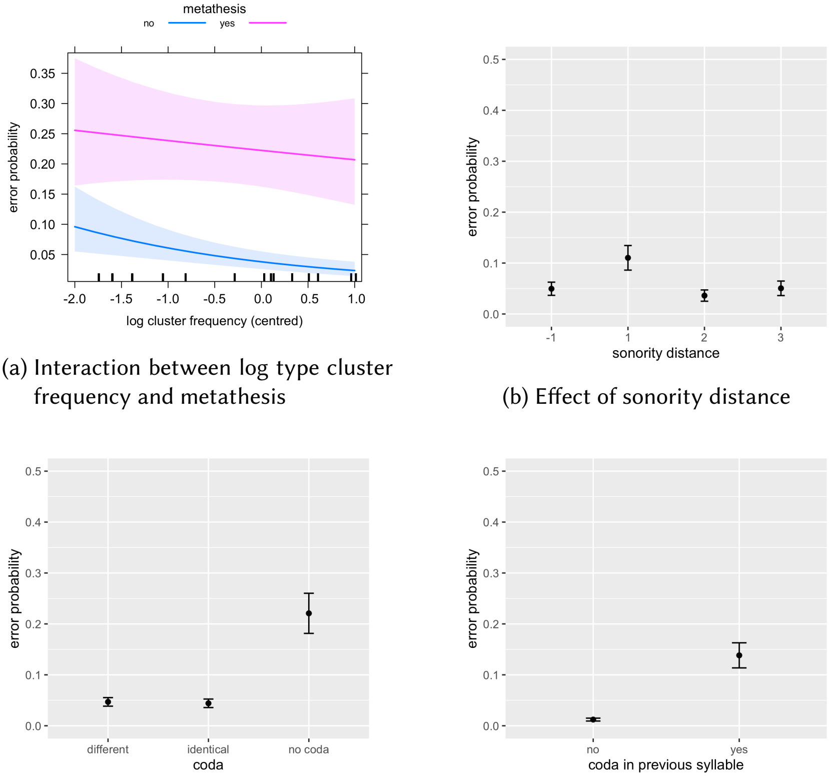
Table 8.3.: Model output of the best-fitting model (complete data set)
Significance codes: *** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1
Formula: `error ~logFreq*metathesis + son.dist + complex + coda.id + (logFreq*metathesis + son.dist|subjID) + (1|onset.targ/stimulus)`

The frequency effect was therefore rather unstable but nevertheless in the direction predicted by usage-based theory: the higher the frequency of a consonant cluster, the lower its error probability (see Figure 8.3a). As the interaction with metathesis shows, this effect was far more pronounced for non-metathesis pairs. For metathesis pairs, the frequency effect was greatly attenuated. The main effect of metathesis was inhibitory, which means that, in cluster pairs with a metathesis, error probability was greatly increased. This was the strongest effect found in the experiment. The effect of sonority distance was only significant for clusters with a sonority distance of 1 (i.e., the stop–sibilant clusters /tʃ/, /ts/, /ks/, and /ps/, as well as the sibilant–nasal clusters /ʃm/ and /ʃn/). For these clusters, the error probability was significantly higher than the average over all groups (see Figure 8.3b). For clusters with a sonority distance of 2, the error rate was marginally significantly lower than the grand mean. The effect of coda identity was significant for all levels: stimuli with identical and non-identical codas had significantly lower estimates and stimuli without a coda significantly higher estimates than the grand mean. However, as this factor was simply added to the model as a control variable because coda conditions were not equally distributed over onset pairs, it will not be discussed further.

In the model containing the binary sonority predictor, SSP violation did not show a significant effect but a trend for SSP-violating clusters to have slightly lower error rates. The results of the other predictors were very similar. Model comparison showed that the model featuring sonority distance provides a better fit to the data than the one featuring SSP violation (AIC: 12808 vs. 12870). Therefore, the binary model will not be discussed in detail here; the model summary and a plot of the sonority effect can be found in the Table B.2 in Appendix B.

When modelling the reduced data set that excluded repetition errors, the effects remained largely the same. The model summary for the reduced data set can be found in Table 8.4.

In the model predicting internal substitutions, both fixed effects were significant, but their direction was opposite to that predicted: the greater the frequency difference between the clusters of a pair (with positive values indicating that the partner cluster is more frequent and negative values indicating that the target cluster is more frequent), the fewer internal substitutions occurred. Likewise, the greater the sonority improvement resulting from a substitution, the fewer substitutions occurred. Table 8.5 shows the model output.



(a) Interaction between log type cluster frequency and metathesis

(b) Effect of sonority distance

(c) Effect of coda

(d) Effect of complex onset (coda in previous syllable)

Figure 8.3.: Significant effects in the tongue twister experiment

predictor	estimate	SE	z value	
(Intercept)	-2.84942	0.14491	-19.663	***
log cluster frequency (type)	-0.22257	0.08979	-2.479	*
cons. metathesis	-0.82535	0.07679	-10.749	***
sonority distance −1	-0.08191	0.15907	-0.515	
sonority distance 1	0.59105	0.13990	4.225	***
sonority distance 2	-0.34814	0.18242	-1.908	.
coda in prev.	-0.90528	0.07200	-12.573	***
coda identical no	-0.37396	0.06526	-5.730	***
coda identical yes	-0.41627	0.07135	-5.834	***
log cluster freq × metathesis	-0.16122	0.04942	-3.262	**

Table 8.4.: Model output of the best-fitting model (data subset excluding repetition errors)
Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$
Formula: `error ~logFreq*metathesis + son.dist + complex + coda.id + (logFreq*metathesis + son.dist|subjID) + (1|onset.targ/stimulus)`

predictor	estimate	SE	z value	
(Intercept)	-4.0360	0.4819	-8.376	***
log frequency difference	-0.8387	0.2035	-4.122	***
sonority improvement	-0.3344	0.1115	-2.998	**

Table 8.5.: Model output of the model predicting internal substitution rate
Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$
Formula: `part.prod ~logFreqDiff + son.improve + (logFreqDiff + son.improve|subj.no) + (1|ons.targ/syllable)`

8.5.2. Comparison of targets and produced onsets

The frequency comparison of target and produced clusters in error productions showed that a cluster was replaced by a higher-frequency cluster more than twice as often as by a lower-frequency cluster (see Table 8.6). There was, however, a big difference between LF and HF targets: LF targets were replaced by a higher-frequency cluster in 78.5%

of all error productions, while HF clusters were replaced by clusters of even higher frequency in only 34.5% of error productions.

In the case of sonority profiles, however, the situation was reversed: most substitutions that led to a change in sonority distance deteriorate the sonority profile of the cluster. Similar to the frequency comparison, there was massive variation depending on the initial value of the target cluster: for SSP-violating clusters, 90.3% of substitutions improved sonority distance, whereas for SSP-conforming clusters, only 4.5% did.

	no. obs.
$\text{freq}_{\text{target}} > \text{freq}_{\text{prod}}$	489
$\text{freq}_{\text{prod}} > \text{freq}_{\text{target}}$	1001
$\text{son. dist.}_{\text{target}} > \text{son. dist.}_{\text{prod}}$	864
$\text{son. dist.}_{\text{prod}} > \text{son. dist.}_{\text{target}}$	332
$\text{son. dist.}_{\text{prod}} = \text{son. dist.}_{\text{target}}$	269

Table 8.6.: Properties of target and produced clusters in error productions

8.6. Discussion

First of all, the results of the present study replicated a number of observations from past research on speech production, the most general one being that speech errors do not occur at random but are facilitated under certain circumstances more than others. It was the primary aim of this study to contribute to defining which circumstances lead to more speech errors and which help to make linguistic items more error-resistant.

One effect that has been reported in numerous studies is the addition bias, in which phonemes are added rather than deleted in speech errors. This has been replicated in the present study: consonant additions outnumbered consonant deletions by a factor of 2.4. In most cases, they showed the execution of two competing speech plans. This can be witnessed in productions like [tksy:l] (for /ksy:l/ in the stimulus pair /tsy:l ksy:l/) with articulation of both the target and the competitor phoneme from the partner syllable, which means that the two alternative speech plans are simply executed in succession. This occurred most often in metathesis stimuli (resulting in onsets like [sks] or [psp]), which shows the increased competition between the consonants involved in these stimuli (see below). The same trend became apparent at a subsegmental level. Both the auditory impression and the spectrogram indicated that, in productions like [ʃ͡flʏt] (in target /ʃlʏt flœp/), there were constrictions in two different places, that is, one phoneme displayed gestures from both the target and the competitor phoneme. This "gestural intrusion bias" (i.e., both the target gesture and an intruding gesture were produced in parallel) has been observed in articulatory studies of slips of the tongue (Pouplier, 2008). Although the present study does not use fine-grained articulatory measures, close inspection of the spectrograms suggests that the gestural intrusion bias is present in the data on top of an addition bias on the segmental level.

For the segmental level, Stemberger (1991; p. 161) explained the addition bias as follows: when an "overt phonological element" competes with a zero element in another word (target), the element has some amount of activation and wins the competition because there is no element in the other word to inhibit it—in Stemberger's words: "*nothing* has no activation level and thus no inhibition". This explanation can also be applied to the subsegmental level: no gesture at a given place of articulation competes with a gesture (e.g., a bilabial constriction) and the latter wins over this *nothing*. At the same time, a gesture at a different place of articulation (e.g., a postalveolar constriction) is executed as appropriate for the target. Although this explanation is appealing, both for the addition of segments and of gestures, it must be refined. Any interactive activation model that builds on this logic needs to include a mechanism that allows zero segments to inhibit activated segments, otherwise additions of phonemes present in the environment would be inflated. The same applies to the subsegmental level. In contrast, Dell's (1986) Spreading Activation Theory predicts null elements to be relatively strong, due to their high frequency. This is also unlikely, and the prevalence of addition errors both in the present study and in previous ones are in conflict with it. Most likely, the truth lies somewhere in between these two extremes.

Another common observation is that errors are often repeated in settings like those in the present experiment. For example, in an experiment similar to this, but using sequences of four real words instead of two pseudowords, Nooteboom and Quené (2015; p. 71) observe "quite some hysteresis": once a particular slip had occurred, it tended to persevere for the remaining repetitions of the stimulus. A similar, although slightly weaker, repetition effect could be observed in the data from the present experiment. Forty-five percent of all errors in the experiment were repetition errors (1352 cases), while recoveries (i.e., a correct production after an error in the same trial) were less common (1137 cases). Since these errors might not follow the same principles as independent errors, another regression model excluding repetition errors was fitted. However, all of the effects remained unchanged (see Tables 8.3 and 8.4).

Not only the presence of a slip before the current production but also the number of preceding productions has an influence on the occurrence of speech errors. Slis and van Lieshout (2016) discarded the first and last repetitions of each trial because, based on previous findings, they assumed that these would behave differently from the other productions. Indeed, the first productions in a trial had a significantly lower error rate than later repetitions in the present experiment, too, χ^2 (3, N = 21758) = 539.85, $p < 0.001$ (Figure 8.4, cf. also Wilshire (1999) for a similar result). The reason for this difference is probably a combination of the repetition effect and an increased cognitive load for later productions, which makes a *first* error more likely in the later productions of a trial. However, the first productions were not discarded here because valuable information can be drawn from them as well.

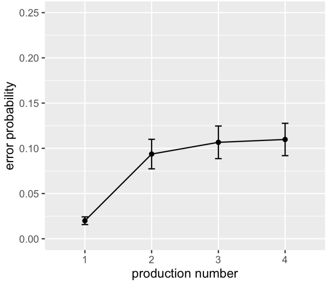


Figure 8.4.: Error probabilities over production number within a trial

The following sections will give more detailed analyses of the main effects.

8.6.1. Frequency

The effect of cluster frequency found in the experiment is not stable over the models fitted, but when present, it supports the hypothesis made in Section 8.3 that production accuracy will be better for HF clusters. The interaction with consonant metathesis indicates that this effect is larger for stimuli with no metathesis.

This is in line with many of the studies reviewed in Section 8.2.1, but is at odds with Santiago et al. (2007), who found an anti-frequency effect on several linguistic levels. These discrepancies might be due to differences in the methods applied. Santiago et al. (2007) analysed a corpus of naturally occurring speech errors in which the environment of all target phonemes was very varied. Although the vast majority of errors in their study was contextual, there was no specific kind of error that was primed for. This is very different from the present experiment, in which competition between two clusters was induced in every trial and the stress level was much higher than in conversation. Moreover, their material was comprised of words, while pseudowords were used in the present study. It has been observed that words and nonwords/pseudowords behave differently when it comes to frequency effects (Vitevitch & Luce, 1998) because processing of the former primarily addresses the lexical level with competition between lexical nodes, while processing of the latter is governed by the sublexical level, in which the higher frequency of sublexical items leads to higher activation of the respective nodes and thus facilitation. Although this observation was made for speech perception, the deviation in results between the present study and that of Santiago et al. (2007) suggests that it might also be true for speech production.

Stemberger (2004; p. 419) himself predicts that the anti-frequency effect he found in SLIP studies with words will be absent in experiments that use nonce words and assumes that in such cases even a weak frequency effect might be observable. His explanation is that nondefault features, correlated with LF phonemes, must be "well-learned lexically" (to overcome the phonological system's general bias for output of HF features) and when they enter competition with default features in HF phonemes, the "strong mapping from meaning to sound" causes them to replace default features in words more often than vice versa. Since nonce words do not have any meaning-to-sound mapping, the nondefault forms are less likely to win the competition here and the general bias towards HF structures becomes visible. Thus, two different explanations have been given for the reversed role of sublexical frequencies in word and nonword processing: while Stemberger (2004) postulated that the decisive step is the learning of sound-meaning combinations for words but not nonwords, Vitevitch and Luce (1998) assumed that the effect arises during processing itself because different levels of processing dominate for words and nonwords.

In the present experiment, consonant cluster frequencies indeed show the facilitative effect predicted by Stemberger (2004) for phoneme frequencies.[21] The higher the frequency of a cluster, the lower its error rate. This shows that, once all effects are eliminated on the lexical level, the facilitative role of sublexical frequencies in speech production can surface. However, in everyday speech production, it is often overridden by stronger effects on the lexical level.

Even on the sublexical level, other effects can render the frequency effect hard to discern (cf. also Vitevitch & Luce, 2005). In this experiment, the frequency effect was partly obscured by the metathesis effect. For future experiments, it would be desirable to have more pairs with diverging frequency that do not have the added difficulty of consonant metathesis. Moreover, an influence of natural class seems to override the frequency effect: the most obvious result of the experiment is that stop–sibilant clusters cause production problems and easily lead to slips of the tongue (also in non-metathesis pairs, see Figure 8.1b). In the cluster pairs that do not contain any of those clusters, error rates are consistently higher for the cluster of lower frequency, with the only exception of /tr/–/kr/, which have very similar frequencies in Elexiko. The role of cluster class will be discussed in more detail in Section 8.6.5.

For now, it is important to highlight the fact that there is a facilitative effect of consonant cluster frequencies on error rates but that it is very fragile. Both its existence and its instability are largely in accordance with findings on the frequency effects of other sublexical units. As discussed in Section 8.2.1, frequency effects do not always surface and sometimes are only visible when examined with very sensitive measures, such as reaction times. It is therefore worth noting that an effect of cluster frequency was visible at all, even when applying a coarse dependent variable, such as repetition accuracy.

Having established that, all other things being equal, LF clusters are more error-prone than HF clusters, the consonant clusters that result from a production error will now be discussed. It was hypothesised

[21] The effect of phoneme frequency could not be determined due to high correlation with cluster frequency.

that slips would more often result in HF clusters than LF clusters. In other words: HF clusters attract responses. The reasoning behind this is that HF clusters present strong competitors for LF targets and, in noisy situations, can win the competition. The comparison of substitution errors resulting in higher- vs. lower frequency clusters shown in Table 8.6 supports this hypothesis: consonant clusters were replaced by higher-frequency clusters more than twice as often as by lower-frequency clusters. The production of target /ps/ as [ts] is an example: even in the absence of direct syntagmatic competition (the two are not paired in a stimulus), the LF target /ps/ is produced as HF [ts] 60 times. The reverse is not true: target /ts/ was produced as [ps] only twice.[22] This replicates a finding by Motley and Baars (1975) that more errors occur when the intruding phoneme is more probable than the target. While their analysis is based on transitional probabilities, the present study suggests that this principle also applies if the resulting sequence of two phonemes is more probable than the target sequence in terms of cluster frequency. Of course, it is also plausible from a theoretical perspective: if LF structures present the speaker with difficulties, it makes sense that the system resorts to an easier structure in order to overcome the problems rather than one of a similar level of difficulty as the target. When analysing this tendency separately for LF and HF clusters, it becomes clear that LF clusters show a much stronger tendency to be replaced by HF clusters than HF clusters do. Almost 80% of LF clusters were substituted with higher-frequency clusters, compared to only around 35% of HF clusters. This replicates a finding by Edwards et al. (2004), who observed that LF biphones were substituted with HF biphones in all error productions, while HF biphones were substituted

[22] It has to be noted, however, that /t/ replacing a competing stop in /ks/ and /ps/ clusters could also be explained independent of frequency, namely in terms of markedness. Romani et al. (2017) also report /t/ as the segment used most often as a replacement in aphasic errors and/or error occurring during L1 acquisition in simple onsets, which can be ascribed to the phoneme's unmarked status.

with higher-frequency biphones in only 50% of errors. This large discrepancy cannot be explained by the mere odds of the availability of more higher-frequency clusters as substitutions for LF clusters than for HF clusters. Rather, it follows from the two frequency principles: firstly, the higher error-proneness of LF structures and secondly, the tendency to resort to HF structures in cases of processing difficulties.

Taken together, the high error rates on LF clusters and the high rate of HF outcomes suggest that, in situations of strong direct competition between two clusters, such as those created in this experiment, the magnitude of the frequency difference between the two competitors will have an influence on the degree of competition. Specifically, the greater the frequency difference between the two competitors, the greater the incentive to substitute the LF cluster with the HF cluster (and the smaller the incentive to substitute the HF with the LF cluster). This should show in a relatively higher rate of internal substitutions in favour of the HF cluster for pairs with a large frequency difference. The partner cluster substitution regression model was run to test this. Since the frequency difference measure was directional (with positive values denoting that the partner cluster is more frequent than the target and negative values denoting that the target is more frequent), the prediction was that substitution rates would increase with an increase in frequency difference.

As noted earlier, the results of the regression model contradicted this prediction. The greater the frequency difference between the competitors (i.e., the higher the incentive for an internal substitution in the case of an LF target), the lower the rate of internal substitutions. This outcome is highly surprising given that the other predictions—more errors on LF clusters and more HF clusters as error outcomes—were supported by the data. These results also stand in strong contrast to the impression of the raw error rates visualised in Figure 8.1b, in which the clusters with the greatest difference generally seemed to have the highest error rates, albeit with a few exceptions. The discrepancy between the results of the separate analyses, on the one hand, and of the model predicting internal substitutions directly from the frequency difference in the clusters of a stimulus, on the other hand, might be due to the relatively low overall rate of internal substitutions (39.0% of all errors). However, the rate of internal substitutions is greater for metathesis pairs (45%), which had proven so influential in terms of error rates, than for non-metathesis pairs (29%). If there is a confounding factor for the metathesis pairs, this would distort the general pattern of substitutions and might have reversed the general picture. As will be discussed later (Section 8.6.5), there is indeed a confounding factor in the metathesis pairs: for them, the cluster's structural makeup is the most relevant influence on error direction and most likely caused the unexpected results for the predictor *frequency difference*.

The regression was therefore repeated post-hoc with a reduced data set that excluded the metathesis pairs. The model summary can be found in Table B.5 in Appendix B. With the exclusion of the metathesis pairs, there was indeed no effect of a high frequency difference leading to fewer internal substitutions. Instead, there was a non-significant trend in the opposite, expected, direction: the higher the frequency difference between the partners, the more frequently the HF partner was inserted.

In general, the hypotheses that HF clusters are stronger and, therefore, both more error-resistant and more likely to be the outcome of a speech error are supported by the data. Noteworthy exceptions for individual clusters will now be examined. The unexpectedly high error rate of /ts/ is probably a carry-over effect due to its pairing with /ks/, which was very difficult to produce. In general, an error increased the probability of further errors in the same trial, also in the partner cluster. In contrast, /sl/, a non-native cluster with a very low frequency, has a remarkably low error rate similar to that of /ʃm/; and /sp/, the cluster with the lowest CELEX type frequency, has a lower error rate than /ks/, /ps/, and /tʃ/. It can be speculated that the CELEX frequency in this case is simply not accurate: WebCELEX lists only six words beginning with /sp/ (*Spin*, *Spiritus*, *Spirituskocher*, *Splen*, *Spot* and *Spotlight*) and ignores many common English loan words, like *Spam*, *Special*, *Space*, *Speed*, *Spot*, *Splatter*, etc., as well as numerous compounds that contain them. Continued contact with English has led to a growing number of loan words so that the numbers in CELEX may well be outdated. For the Greek-origin onsets /ks/ and /ps/, on the other hand, there are virtually no new lexemes, so the CELEX numbers are more reliable. A similar argument can be made for /sl/, with CELEX lacking words like *Slash*, *Slot*, *Slow-* (as in *Slow-Food* or *Slow-Motion*), *Slideshow*, *Slapstick*, *Slackline*, etc. An alternative explanation would be that /sl/ and /sp/ benefit from the high frequency of clusters with a similar make-up in terms of natural classes (e.g., /ʃl/ for /sl/ and /ʃp/ for /sp/). To explore this possibility, generalised cluster frequency (summed over natural classes) was added post-hoc to the model described in Table 8.3. The results showed that the effect of generalised cluster frequency was significant (p < .001, see Table B.3 in Appendix B) but inhibitory, which indicates that a high class frequency adds to the competition rather than facilitating production. This means that generalised cluster frequency cannot explain the exceptions listed above. A third potential explanation, which can account for all ostensible exceptions in a uniform way, will be discussed in Section 8.6.5.

The existence of frequency effects of sublexical units also has implications for the interpretation of categorical phonotactic effects. If the human speech production system is biased towards the output of frequent phoneme strings, as has been shown here, then there is no need to postulate a separate mechanism that filters output in terms of phonotactic rules. Phonotactically illegal strings are simply extreme cases of low frequency, which the production system accordingly has a strong bias against. Dell et al. (1993a) have demonstrated this in computer simulations of English CVC word production, in which the parallel distributed processing model did not contain any explicit phonotactic rules but derived a very strong tendency for phonotactically legal output (83–100%) solely on the basis of adequate input vocabulary and feedback mechanisms concerning the sequential progress.

Another interesting aspect about the frequency effect found here is that it emerged at a high speaking rate. According to the Spreading Activation theory, the frequency bias is stronger at slower speaking rates. Since the speaking rate was held constant in the present experiment, this prediction could not be falsified; if it holds true, the clear frequency effect found at such a high speaking rate in the present experiment would be remarkable.

8.6.2. Neighbourhood frequency

There was no significant effect of neighbourhood frequency. In contrast, previous studies have often found either facilitating or inhibitory effects of phonological neighbourhoods on the lexical level. In the perception experiments reported in Chapters 5 and 6, clear inhibitory effects of cluster neighbourhood frequencies emerged, which shows that the sublexical neighbourhood measure used here—the summed frequencies of all initial clusters that differ from the target in one phonological feature—is a valid implementation of neighbourhood structures on the sublexical level. There are two explanations for the absence of a neighbourhood effect. Firstly, it is possible that there was both a facilitating and an inhibitory effect of neighbourhood frequency, as Sadat et al. (2014) suggested, but that the two opposing forces levelled each other out, so that none of the effects could reach significance. Sadat et al. (2014) noted that facilitating effects can emerge in instances of disrupted speech production, whereas neighbourhood effects are usually inhibitory in unimpaired speech production. While unimpaired speakers were tested in the present experiment, the production task was far more demanding than normal speech production. This situation might have mimicked impaired speech production. For example, Meffert et al.

(2011) show how aphasic behaviour can be induced in healthy speakers by applying production tasks with high cognitive demands. Hence, the inhibitory effect of dense, HF neighbourhoods typically found in healthy speech production may have emerged concurrently with facilitative effects caused by the demanding task, so that the two cancelled each other out.

The second possibility is that, although cluster neighbourhood frequency proved to be an adequate measure in the perception experiments, it might not have been differentiated enough for the production task. Recall that Stemberger (2004) argued that overall neighbourhood size does not affect speech production and instead suggested dividing phonological neighbours into friends and enemies. In this case, too, opposing forces can be said to have levelled each other out. Regardless of which of the two explanations holds, it can be stated that the relationships between phonological representations in the mental lexicon are complex and the manifestation of an effect depends on many factors. In the task at hand, cluster neighbourhood frequencies were not a suitable measure to show a visible effect on production accuracy.

8.6.3. Sonority

Throughout most of this dissertation, a potential processing difference between SSP-adhering and SSP-violating clusters has been explored. This binary division did not yield a significant effect in the logistic mixed effects regression models; but since the literature on speech production and acquisition mostly reported effects of sonority distance, this finer measure was also used in a separate model. When this measure was included, a sonority effect became visible: clusters with a sonority distance of 1 had significantly higher error rates than all other clusters. Obviously, the effect is not monotonous. This is at odds with phonological theory, which predicts that error rates should be highest for clusters with a sonority distance of -1 (i.e., clusters that violate the

SSP) and then drop steeply towards a sonority distance of 1 and more gradually—if at all—afterwards. This is because violations of the SSP are expected to be most problematic. As reviewed in Section 8.2.2, SSP-violating clusters have been shown to be error-prone, at least for some populations (mostly apraxic and aphasic speakers, and partly in children during L1 acquisition). However, according to the SDP, there are well-formedness differences even within the group of SSP-conforming clusters. The steeper the rise in sonority syllable-initially, the more well-formed the syllable. Therefore, one might even expect a slight decrease in error rates from clusters with a distance of 1 to 2 and 3. The data clearly show that this is not the case. The group of clusters with the highest error rate is composed of stop–sibilant clusters and sibilant–nasal clusters. Figure 8.1a (p. 276) shows that it is the group of stop–sibilant clusters that causes this effect, while /ʃm/ and /ʃn/ have low error rates. Although their sonority profile is not optimal, there is no reason from a sonority-theoretical perspective why they should cause more problems than their reversed counterparts, which violate the SSP. The difference in error rates thus cannot be satisfactorily explained by sonority theory and the hypothesis that sonority sequencing guides pseudoword production in healthy adult speakers was not supported.

In terms of error outcomes, sonority theory predicts—and it was hypothesised—that errors will improve sonority profiles, as can often be observed in apraxic speech errors. However, as Table 8.6 on p. 282 shows, the opposite is the case in the data from the present experiment: in the majority of cases, the sonority profile was deteriorated by the speech error; only in a minority of cases was it improved or maintained. The outcome analysis therefore contradicts sonority predictions just as strongly as the analysis of error rates.

As in the frequency analysis, it was also analysed whether the difference in sonority distance values within a pair had an influence on substitution rates. The substitution error model revealed that the larger the sonority improvement would be in the case of an internal substitution, the fewer internal substitutions occurred.

Summing up, sonority sequencing cannot account for the patterns found in the error data from the experiment. This is in line with previous research that finds sonority effects primarily in children and speech-impaired populations and for illegal clusters. It suggests that, even under increased processing pressure, healthy adults are not influenced by sonority sequencing. In contrast to children and L2 learners, the overlearning of their native phonotactic system seems to have desensitised them.

Another potential explanation for the lack of an interpretable effect is that sonority is not merely an articulatory phenomenon. If this were the case, effects of sonority sequencing should arise in speeded production tasks, as mentioned in Section 8.2.2. It can be speculated that sonority effects found in aphasic and apraxic patients are located at the level of the phonological lexicon, as suggested by Kohn et al. (1998) and Romani et al. (2011), and, consequently, did not emerge in this experiment because the phonological lexicon did not play a role.

The hypotheses that SSP-violating clusters will have higher error rates and that errors will result mostly in SSP-conforming clusters are not supported by the data, whereas the hypothesis that a cluster's frequency will have a larger influence on its production accuracy that its sonority status is supported by the present results.

8.6.4. Consonant metathesis and competition

The strongest effect in the tongue twister experiment was that of consonant metathesis: the cluster pairs /ʃt–tʃ/, /sk–ks/, and /sp–ps/ had significantly higher error rates than all other pairs. This is very obvious when examining the clusters that occur both in a metathesis pair and in a non-metathesis pair. For all three clusters, error rates were significantly higher in the metathesis pair than in the non-metathesis pair (statistics for Welch's t-test can be found in Table B.6 in Appendix B). For example, the error rate for /ks/ was as low as 20% in the pair /ks–ts/, but rose to 50% when competing with /sk/. This shows that it is not only a cluster's inherent difficulty that determines the error probability but also the degree of competition it receives during the planning of an utterance. This competition is greatest when a consonant cluster alternates with its reversed counterpart. Note that, in contrast, /fl/ had equally low error rates of 3.0% and 3.5% when competing with /sl/ and /ʃl/, respectively.

A likely explanation based on a connectionist model of speech production is that both consonants of the onset clusters are strongly activated because they occur (at least) twice in the planned utterance. This increases the competition between them and can lead to the wrong segment being produced in situations of increased cognitive load. In terms of Dell's Spreading Activation theory, there are two possible causes for this error: On the one hand, a consonant in a metathesis pair receives activation as the current node on the phoneme level simultaneous to anticipatory activation from the cluster level, on which the following syllable is already prepared, leading to its activation level reaching threshold and it being erroneously inserted into the syllable. On the other hand, a cluster that is not the current node on the cluster level any more receives feedback from its component phonemes on the phoneme level (which are the current nodes on the phoneme level due to the fact that they occur in the following syllable), causing the insertion rule to err. This is what Dell (1986) calls "confusion between levels" (cf. Section 7.5).

What makes the two competing consonants so confusable is that they appear in the same position in the course of the utterance. Not only are they both *onset* consonants and hence specified for the same syllabic position, which is known to be a precondition for contextual substitutions, but they also appear in the same position *within* the onset in the course of the utterance. The fact that this adds to their confusability in speech production has very interesting implications for linguistic theory. It supports Stemberger and Treiman's (1986) hypothesis of separate skeletal positions for C1 and C2 (cf. Section 8.2.5) because, if consonants are specified for C1 vs. C2 position, then there will be contradictory information as to which skeletal position each of the consonants belongs to in the cases of metathesis pairs. There might be an inhibition mechanism that can prevent a consonant that is also specified for C2 position from being uttered in C1 position and vice versa. The stimulus sequence /ksaːn skaːn/ can serve as an example. During the planning of the utterance, both /s/ and /k/ will be highly activated because each of them receives activation both from /ks/ and from /sk/ on the cluster level, but each of them will be specified for both C1 and C2 position. Therefore, during the planning of the first syllable, /k/—due to its additional tag as C2—might receive slight inhibition for insertion in C1 position and /s/—which is also specified for C1 and strongly activated—will be presented as a viable alternative. This means the metathesis pairs are difficult because, during the production of each of the onset consonants, this consonant receives slight inhibition due to its being specified for the other consonantal position in the onset, too, while at the same time a highly activated competitor is available for the same position. In non-metathesis pairs, in contrast, there would not be any inhibition for insertion in a certain position (because a given consonant is only tagged for one position), only competition between the two consonants for one of the positions. For example, in /ʃt–ʃn/ stimuli, /t/ and /n/ compete for C2 position in both syllables, but each of them is unambiguously labelled for that position, so that insertion of the target is not inhibited.

It is noteworthy that such strong competition was induced between the onset consonants in the metathesis pairs, while there was no increased competition between identical consonants in onset and coda position (e.g., /n/ or /t/ in the stimulus /ʃtaːn ʃnaːt/). This observation is in line with previous research that shows that contextual errors involve segments in the same syllable position (e.g., Fromkin, 1971; MacKay, 1978). It is also well accounted for by Dell's Spreading Activation theory, in which consonants are specified for syllable position and only an item with the correct tag for a given slot can be inserted into that slot (see point 3 in the description on p. 221).

8.6.5. Special cases: Sibilant–stop and stop–sibilant clusters

When inspecting the errors among the metathesis pairs more closely, a clear pattern of internal errors emerges: in all cases, the stop–sibilant cluster is substituted by the sibilant–stop cluster substantially more often than the other way around. For example, by far the highest number of internal substitutions—and in fact errors in general—is attributable to /ks/ > /sk/, as can be seen in the confusion matrix, Table B.7 in the Appendix. The pattern is similar, although less extreme, in the other metathesis pairs, which suggests that it is the structure of the clusters (i.e., their composition in terms of natural classes) that determines their strength. This can, in part, be explained by the metathesis effect discussed above, but stop–sibilant > sibilant–stop repairs are not restricted to metathesis pairs. For example, target /ks/ was produced as [sk] 64 times in the stimulus pair /ts–ks/. This indicates a general tendency to replace the problematic stop–sibilant clusters with their reversed counterparts.

On a more general scale, the plots of error rates across clusters (Figure 8.1a, p. 276) and false positives across clusters (Figure 8.2, p. 277) confirm this picture: on the one hand, the three stop–sibilant clusters have the highest error rates, and on the other hand, sibilant–stop clusters have the highest rates of false positives. In other words: sibilant–stop clusters appear to be strong in that they often act as intruders (even in non-contextual errors), while stop–sibilant clusters appear to be weak in that they are the most error-prone clusters. The fact that stop–sibilant clusters, in addition to having the highest error rates, also had the second highest rates of false positives can be explained in terms of exchange errors: if, in a metathesis pair, a slip occurred on a stop–sibilant cluster (producing a sibilant–stop cluster instead), participants were more likely to restore the alternation of the two cluster types and produce a stop–sibilant cluster for the next sibilant–stop target.

The diverging strengths of the two kinds of clusters can also account for the discrepancies with respect to the role of cluster frequencies noted above. While there was a general tendency for HF clusters to resist errors and to act as intruders, this was not reflected in a higher rate of LF clusters being replaced specifically by their HF partners. It turns out that this was due to an opposite effect in metathesis pairs. In them, the structure in terms of natural classes is the most decisive factor in determining the direction of substitutions.

The special status of sibilant–stop clusters, both in terms of their distribution and their role in speech processing, has received some scholarly attention over the past decades (Dziubalska-Kołaczyk, 2015; Goad, 2011; Morelli, 1999). In spite of the fact that they violate the SSP, they are relatively common cross-linguistically (Morelli, 1999) and are acquired early on in L1 acquisition (Dziubalska-Kołaczyk, 2015). They also stand out phonetically and articulatorily in that they lack stop aspiration and only have a single glottal gesture (Browman & Goldstein, 1986; Byrd & Choi, 2010). In a speeded naming experiment, they had shorter response latencies than words with singleton /s/ onsets, although latencies had previously been found to increase with the number of phonemes (Kawamoto & Kello, 1999). Numerous accounts have been proposed regarding their structural representation to attempt to resolve the dilemma of such a common class of consonant clusters as violating sonority sequencing. They range from an extra-syllabic position of the sibilant (J. Harris, 1994) and syllabification processes after core syllabification, which is the domain for sonority (Clements, 1990), to such sequences as single, complex segments (Browman & Goldstein, 1986; Fudge, 1969; Selkirk, 1982; but see Treiman, 1986; for counter-evidence) or at least as having high intersegmental cohesiveness (Berg, 1989; Tzakosta, 2009). In fact, they are often mentioned alongside affricates.

In an interpretation of a pre-stop sibilant as extra-syllabic, the cluster's cohesiveness would be very low and the cluster should not behave like a unit. One would therefore not expect it to show frequency effects, either. However, the data from the present experiment showed clear frequency effects for such clusters (compare, for example, error rates of HF clusters /ʃt/ and /ʃp/ with those of LF clusters /sk/ and /sp/). The explanation of extra-syllabicity is therefore at odds with the present findings.

An interpretation of sibilant–stop clusters as having high internal cohesiveness and being more unit-like, on the other hand, would explain their relative strength in the error data from the tongue twister experiment. The more cohesive a cluster is, the more error-resistant it should be. According to Berg (1989), sequences with high cohesiveness are more strongly activated at any given moment during speech production because they are dominated by the same syllable structure node and receive their activation from it in parallel. If one assumes that sibilant–stop clusters have a high degree of cohesiveness and hence also high activation, their high rate of false positives is less surprising. What remains to be explained, however, is why the reversed structure, stop–sibilant clusters, is so problematic in production. The difficulties they caused in the experiment cannot be reduced to the metathesis effect, although it contributed a great deal to them. However, /ps/ also had a high error rate in /ps–pl/ stimuli, as did /ks/ in /ks–ts/ stimuli.

At this point, it is interesting to take a look at Tzakosta (2009), who makes a four-way distinction between complex onsets. Based on the data from cluster reduction during Greek L1 acquisition, she discriminates between 1) true clusters (e.g., stop–liquid), 2) sibilant–stop clusters, 3) stop–sibilant clusters, and 4) affricates, to which she ascribes different degrees of *cluster coherence*[23] (listed here in order of ascending coherence). The two cluster groups that stick out in the present experiment are contrasted with true clusters in her analysis (along with affricates). The most common repair strategy for true clusters is reduction which follows the principle of sonority optimisation, that is, the less sonorous consonant in C1 position is preserved, leading to a steep rise in sonority in the transition to the vowel. When compared to the data from the adult speakers in the present experiment (see confusion matrix, Table B.7 in Appendix B), one can see an overwhelming tendency for true clusters to follow the same principle. Both stop–liquid and fricative–liquid (including sibilants and /f/) clusters follow the pattern of sonority optimisation described by Tzakosta (2009) and also D. K. Ohala (1999): the more sonorous element is deleted more often, thereby creating a steep rise in sonority towards the syllable nucleus. In the four obstruent–liquid clusters that do not involve sibilants, C1 is always preserved, while C2 is deleted relatively often. In the sibilant–liquid clusters, this tendency is almost as strong; /ʃm/ is the only obstruent–sonorant cluster that does not follow this pattern. Here /ʃ/ is deleted more often than /m/.

Both sibilant–stop and stop–sibilant clusters, on the other hand, show the opposite trend: the less sonorous stop is deleted most often,[24] which is a reduction strategy contra sonority principles (although, of course, reduction to the sibilant also constitutes a sonority improvement when compared to the full cluster). In terms of sonority, they thus behave unlike any other cluster.[25]

A high rate of addition errors is also common to both groups, the majority of which are productions of C2–C1–C2 sequences (e.g., [sps]

[23] *Coherence* in her terminology denotes the same property as Berg's *cohesiveness*.

[24] with the exception of target /ʃp/, in which /p/ was never deleted but /ʃ/ three times

[25] The strong resistance of sibilants to deletion is surprising since they are usually found to be very prone to loss in speech errors (Miozzo & Buchwald, 2013; Shattuck-Hufnagel & Klatt, 1979; Stemberger & Treiman, 1986; Tzakosta, 2009) but will not be pursued any further here.

for target /ps/). Subjects often started to produce the wrong onset in metathesis pairs and then fused the two competitors into a more complex (and illegal) cluster. This fusion can be taken as an indication that both competitor clusters have a high level of cohesiveness. If stop–sibilant clusters, like sibilant–stop clusters, form a (relatively) cohesive unit, however, why are they by far the most problematic cluster group? The most straight-forward explanation is that, as a cluster class, they are not native (i.e., they only occur as onsets of loan words) and have a low frequency of use.[26] Interestingly /ts/, which is considered an affricate in German phonology, patterns with /ps/ and /ks/, both in terms of increased error rates (considering its high frequency) and in terms of which element is deleted. It is therefore reasonable to assume that /ts/ does not have a special status as an affricate. Whether this is due to the experimental design in which it contrasted with true clusters of the same make-up (and was even paired with /ks/ in the same stimuli) or applies to all speech production can, at this point, not be determined with certainty.

The results of the current experiment add further evidence to a growing body of research that shows that sibilant–stop clusters and stop–sibilant clusters have a special status in a number of languages and behave differently from other clusters in speech production as well. This means that, in addition to language-specific factors like cluster frequencies, a structural and potentially universal component does play a role in consonant cluster production: sibilant–stop clusters are easier to produce than stop–fricative clusters. However, it is not related to SSP violations but rather seems diametrically opposed to the concept of simplification through steady sonority growth within the syllable.

[26] Note also that, in terms of error rates, the positions of stop–sibilant vs. sibilant–stop are reversed in the present data when compared to the hierarchy set up by Tzakosta (2009). Sibilant–stop clusters are stronger (i.e., have lower error rates and higher rates of false positives) than stop–sibilant clusters. A possible explanation for this is that stop–sibilant clusters are native to Greek—the language studied by Tzakosta—, while in German they only occur in loan words.

What constitutes this component is still open for debate. The results of this experiment support an interpretation in terms of high intersegmental cohesiveness. The only thing that goes against this explanation is the fact that these two cluster types are split up by speech errors more often than the other clusters; the opposite would be expected for highly cohesive units. The deviation from error rates that would be expected on the basis of their frequencies is also reminiscent of Edwards et al.'s (2004) finding that CV sequences are produced more accurately than their frequencies would imply. This hints towards the more general importance of skeletal structure when making predictions based on biphone frequencies.

If these special cases are left aside, though, an influence of sonority on error patterns is visible: in deletion errors in "normal" clusters, it is almost always the more sonorous consonant that is deleted, which creates a steep rise in sonority at the onset. The expected effect of higher error rates on SSP-violating clusters and their repair as SSP-conforming clusters does not emerge, however, unless stops and fricatives are assigned the same sonority value and both cluster types discussed here are thus considered to be plateau clusters. In that case, however, the substantially higher error rates of stop–sibilant clusters (as opposed to sibilant–stop clusters) cannot be explained.

8.6.6. Grain sizes and consonant clusters as linguistic units

In Section 8.2.1, it was discussed which linguistic units display frequency effects in speech production and under what circumstances. It was also mentioned that for frequency effects to occur for a given linguistic unit of representation, it is vital that this unit plays a role in the language under investigation. By the same logic, frequency effects for a linguistic unit are often taken as evidence for that unit's psychological reality.[27] If consonant clusters are processed as holistic units, one would expect a frequency effect on that level. The data from this experiment revealed that such an effect does, indeed, exist. It will now be discussed how consonant cluster frequencies behaved relative to frequencies of other linguistic units in the present experiment and what this means for their roles in German speech processing. It is important to note, however, that in a process as complex as speech processing, there might not be a primary processing unit but several units that each play a role at some point during the process, in production as well as in perception.

During model fitting, syllable type frequency of the stimuli had to be removed because it did not contribute significantly towards explaining the data, while consonant cluster frequency did (although it was not stable across models). Single segment frequency could not be included in the model because it was too highly correlated with cluster frequency. Instead, a model was fitted in which cluster frequency was replaced by summed phone frequency in order to compare their effects. The model summary can be found in Table B.4 in the Appendix. All of the effects from the model with cluster frequency as a predictor remained the same. When comparing the Akaike information criterion and Bayes information criterion for the two models, it turns out that the phone model explains the data better than the cluster model (AIC: 12,785 vs. 12,808; BIC: 13,104 vs. 13,127). This better performance of the phone model is remarkable since it was not the frequency of a single phone that predicted the error probability of that phone, but rather the summed frequency of the two onset phones predicted the error prob-

[27] But see Wade et al. (2010) for a critical account of this common reasoning and a demonstration of how syllable frequency effects can be derived without explicit reference to the syllable level in speech production, just by taking the (acoustic-phonetic) context that surrounds a segment to be chosen for production into consideration.

ability of the cluster as a whole. It is conceivable that a more exact model of phone frequency would fare even better.

It could now be argued that the effect of cluster frequency observed in the main model is an artefact of its correlation with phone frequency. However, if that were the case, the error probabilities for clusters composed of the same phones should not diverge as much as observed because their summed phone frequency is the same. It must therefore be concluded that cluster frequency does contribute towards explaining the error pattern. Both phone frequency and cluster frequency seem to influence the production of the pseudowords used in this study, with phone frequency serving as a better predictor. Syllable frequency, on the other hand, did not show a significant effect. Romani et al. (2016) obtained similar results when comparing the influences of phoneme and syllable frequencies on aphasic speech errors: phoneme frequency had a facilitating effect on aphasics' speech errors, while syllable frequency showed no or paradoxical effects. It seems, therefore, that for sublexical units, smaller ones are of greater importance to production than larger ones. The fact that the present study used pseudowords as stimuli and unimpaired speakers acted as subjects, while Romani et al.'s study looked at the production of words by aphasic speakers, together raises hope that this finding is generalisable.

The relationship between single segments and consonant clusters as units in speech production deserves further attention. In a study with aphasic and apraxic patients, Jakob (2018) found that patients with phonological impairment (aphasia) produced two of the four initial consonant clusters in the study more accurately than would be expected on the basis of the error rates of the component consonants. This suggests that these patients combine the two consonants to form a unit, which is then processed holistically. This effect was only found for /kl/ and /ʃl/, not for /kn/ and /ʃn/, which showed error rates that corresponded to the combined error rates of their component phones. This finding is interesting because the separation of effects mirrors a frequency border between the clusters. While /kl/ and /ʃl/ have CELEX type frequencies around 330 (token: > 3600), the type frequencies of /kn/ and /ʃn/ are lower than 150 (token: < 800). This could be taken as an indication that only HF clusters are processed as units, while LF clusters[28] show no unit effect and instead behave like a combination of individual consonants for patients with phonological aphasia. (The picture was completely different in the apraxic data, in which three clusters had error rates equal to what would be expected on the basis of the error rates of the component consonants and one cluster, /kn/, had a higher error rate.) It is plausible that this is also true for unimpaired speakers. The pattern is certainly reminiscent of Levelt's syllabary, in which only HF syllables are stored, while LF syllables are assembled online.

However, it is known that for real words, the effect of sublexical frequencies is weaker than that of lexical frequencies and other lexical variables (Andrews, 1992; Vitevitch & Luce, 1999). It is therefore conceivable that motor programmes for larger units, like whole words or even phrases, exist, which can override the effects of sublexical frequencies. When there is no pre-fabricated motor programme available for a word as a whole, as is the case for neologisms and maybe other, infrequent words, speakers may resort to sublexical units and their frequencies for motor facilitation (see also Shuster, 2009; p. 69 for a similar argumentation).

8.6.7. Location of the frequency effect

Frequency effects of sublexical units have often been found at the stage of phonetic encoding (Cholin & Levelt, 2009; Laganaro & Alario, 2006)

[28] It should be noted, however, that in relation to the whole frequency range of German consonant clusters, all four clusters examined in Jakob's study lie somewhere in the middle of the scale. This means that, if frequency differences are responsible to the divergent error rates of those four clusters, there would have to be a sharp split that goes through that central region of the frequency scale.

and sometimes at the intersection between phonological and phonetic encoding (Perret et al., 2014). Furthermore, Munson (2001; p. 790) remarks that "information about pattern frequency is encoded at many levels of linguistic representation" since he found frequency effects of sublexical units on the level of phonological encoding as well as on lower levels. Bose et al. (2007) observed effects of bigram frequency on lower-level speech production processes, namely speech motor performance, and conclude that they are not limited to speech preparation processes, as is traditionally assumed. Likewise, Ziegler and Aichert (2015; p. 36) suggest that the between-gestural bonds they observed in their study of apraxic speakers are "the result of an extensive motor learning process, through which the frequently occurring gestural patterns of a language become particularly entrained, even though they may not conform to ostensible economy principles". They thus identify a frequency effect similar to the one observed here on the motor planning or motor execution level.

It therefore makes sense to ask where exactly the effect of cluster frequency found in the tongue twister experiment is located. The lexical level can be excluded given that it was not addressed by the stimuli and the frequency effect concerns a sublexical unit. The small number of lexical outcomes also suggests that it was not particularly active during the experiment. There were only 188 syllable productions that constituted lexical items and these were distributed over many different syllables with mostly just one occurrence per syllable.

The tongue twister paradigm is well-known as a technique for the study of phonological encoding (Wilshire, 1999), hence the frequency effect is likely to have emerged at this stage. The speaking rate of 144 syllables per minute (which corresponds to 2.4 syllables per second) is also below the rate that has been shown to affect motor articulation (Laver, 1980). Moreover, the fact that only *type* frequencies had a significant effect on error rates supports an interpretation in terms of activation of higher-level units which contain the clusters. If pure practice of motor patterns had been the source of the effect, it should also have been present—and even more pronounced—for token frequencies.

This leaves phonological, phonetic, or motor encoding as possible candidates. With the task applied in the present study, it is not possible to pinpoint exactly at which of these stages the effect of cluster frequencies manifests itself. Phonetic accommodation might shed further light on the dilemma since it is often viewed as a good indication of whether an error occurred at the phonological or the phonetic stage. If the produced sound is accommodated to the new environment, the error can be assumed to have emerged at the phonological level and then integrated during phonetic processing. Errors like /ʃp/ > /p/ are therefore a good indicator: if the stop is produced with a long VOT and aspiration, as is appropriate for a singleton onset stop, it is accommodated to the new phonological context. If it is produced with a short VOT and no aspiration, as is standard in sibilant–stop clusters, it has not been accommodated to the new environment and has most probably occurred after phonological processing. Analysis of a random sample of this kind of reduction errors reveals that most, but not all, of them are phonetically accommodated. This means that the majority of errors seem to have arisen at the level of phonological encoding. However, there are also examples of phoneme deletions without subsequent accommodations of the remaining stop (e.g., /ʃtɔr/ > [dɔʁ]), which most probably originate at the phonetic or motor encoding stage of production. Keeping in mind Munson's (2001) results of biphone frequency effects both on the level of phonological encoding and below, it is quite likely that in the present experiment, too, there are several different sources of the effect.

8.6.8. Implications for models of speech production

What conclusions with respect to models of speech production can be drawn from the results of the present experiment? The absence of a syllable frequency effect does not necessarily exclude the possibility of a mental syllabary. The dependent measure of error rate is probably too coarse to capture the differences between syllables stored as units and syllables assembled on-line. Even though differences between them have been found in processing time (e.g., Cholin et al., 2006; Levelt & Wheeldon, 1994), there is no reason to assume that the on-line-assembly route leads to correct results less often than access to the stored syllable in the syllabary.

That said, the syllable frequency effect observed by Levelt and colleagues can be explained at least equally well by connectionist models, which can also account for the present results. In connectionist frameworks, frequency effects are based on stronger connections between units that are frequently used together. In the present case, the connections between the two consonants of HF clusters would be strengthened. The model would have to be able to account for the absence of a frequency effect on the syllable level and the presence of frequency effects on the biphone (consonant cluster) and phoneme levels, though.

Another feature that is important to implement in speech production models is the increased competition between phonemes in utterances that contain the same phoneme combinations with slight variations in several places (such as the metathesis pairs). One way in which competition can be accounted for is via the decay rates of previously selected nodes.

In general, the way competition is dealt with is a central issue to models of speech production. They also need a mechanism to handle the possibility that speech errors preserve some features of the target phoneme. In cascading-activation models, this is implemented in terms of gradient activation of representations. Activated nodes can pass on their activation to connected nodes before they are selected. Following an analysis of subfeatural speech errors, Frisch and Wright (2002; p. 160) remark that existing models need to be "extended to include competition at the level [of] [...] *articulatory plans of segments*" (emphasis added). The results of the present analysis support this interpretation. Many errors had characteristics of both the target and a competitor phoneme, which can most plausibly be explained by competition between articulatory plans leading to the partial execution of several plans.

The data from the present experiment also replicates previous findings that errors tend to be repeated in repetitions of the same utterance(s). This suggests that subjects did not compute the syllable repetitions from scratch each time but at least partially reused previous speech plans—also erroneous ones—during repetitions (cf. Wilshire, 1999; for a similar assumption, albeit based on a different observation). With regard to the implementation of this process, Wilshire (1999; p. 75) writes:

> Within a simple network model of speech planning, we might propose that the sequential pattern of activity generated across phoneme representations during the initial production of a string can be stored temporarily, and regenerated for any subsequent recitation. This would reduce the amount of new phonological processing that had to be performed.

Summing up, connectionist models of speech production, especially those that feature cascading activation, can account for the main findings of this experiment. Most of the data are in line with Dell's (1986) Spreading Activation theory, although a mechanism that creates subphonemic speech errors would have to be added. Another difference between the present data and the results of Dell's (1986) simulations is that the latter created more deletion errors than substitution errors, while the opposite pattern was found in the present experiment. This dilemma could be solved by the adapted model version, which does not involve null elements. The main findings from the experiment, however,—the facilitating effect of cluster frequency and the inhibitory effect of consonant metathesis—can be accounted for by the Spreading Activation theory.

8.6.9. Summary

The experiment presented here succeeded in eliciting contextual speech errors by introducing competition between similar clusters. In addition to an effect of complexity (operationalised as the number of consecutive phonemes), it demonstrated an effect of consonant cluster frequency, which was unstable across models but, nonetheless, has important implications for our understanding of how previous experience(s) and statistical reckoning are used during on-line cognitive processing. Frequency effects on accuracy are hard to find for the population studied here—adults without language impairment (see literature review). By enforcing competition between two clusters, however, it was possible to make visible an effect of cluster frequency. Clusters of higher frequency tended to have lower error rates and higher rates of false positives, but there are a number of other factors that can distort this pattern. Most notably, some combinations of natural classes are very strong, while others are very error-prone,[29] especially in a metathesis context. This led to a masking of the frequency effect for internal substitutions. The highest number of contextual errors were induced by consonant metathesis; for metathesis pairs, frequency was not the most relevant factor in determining the direction of error asymmetry. In contrast to frequency, sonority sequencing did not show the predicted effect, unless obstruent consonants are treated as a homogeneous group and all obstruent-obstruent clusters, therefore, as ille-

[29] The data on stop–sibilant clusters might also explain the scarcity of such clusters in the German language. Of course, it is difficult to distinguish between cause and effect with certainty, but it seems plausible that stop–sibilant clusters, which caused so many problems for the subjects in this study, are absent in the German language and rare in loan words precisely because they are difficult to pronounce.

gal plateau clusters. In that case, the strongly diverging behaviour of sibilant–stop vs. stop–sibilant clusters cannot be explained, however.

8.6.10. Conclusions and future directions

When comparing the effects of cluster frequency and sonority, frequency makes more reliable predictions of error patterns. Table 8.7 gives an overview of diverging predictions for error rates based on frequency and sonority values along with the observed error rates in a few selected cluster pairs. Only pairs with a clear frequency difference are considered since it is otherwise difficult to determine a clear cut-off point at which error rates should differ significantly. (Moreover, for some of the LF clusters as well as /tr–kr/, CELEX and elexiko disagreed concerning their frequency rankings.) All reported differences in observed values are significant at a .01 level according to χ^2 tests.

cluster pair	frequency prediction	sonority prediction	observed
fl–sl	sl higher	same	sl higher
ts–ks	ks higher	same	ks higher
ʃt–tʃ	tʃ higher	ʃt higher	tʃ higher
ʃt–ʃn	ʃn higher	ʃt higher	ʃn higher
ʃp–ʃm	ʃm higher	ʃp higher	ʃm higher

Table 8.7.: Predictions for comparison of error rates within cluster pairs and observed values

As can be seen from the table, frequency predictions were correct in all pairs that have a large frequency difference (and for which CELEX and elexiko agree on a ranking of the two). The sonority predictions, on the other hand, were always wrong. Hence, the hypothesis that frequency distributions of consonant clusters are more important for their processing in speech production than sonority sequencing is supported by the data.

There are, of course, some limitations as to the generalisability of data from the experiment and the conclusions that can be drawn from them. A study that relies entirely on transcription data must be interpreted with caution (see Frisch & Wright, 2002; Pouplier, Marin, & Kochetov, 2017) because many subphonemic errors cannot be captured by this coarse metric but only by fine articulatory or acoustic data. The inspection of the production data also very clearly showed that many production errors were subphonemic articulation abnormalities that were impossible to accurately analyse with the methods applied here. Special care was taken to annotate subphonemic details and many productions were played numerous times, often with the support of inspection of the spectrogram before a transcription was made. Nonetheless, these data can never capture as many details as articulatory or quantitative auditory analysis can. The subphonemic errors were not considered in the regression model. It would be desirable to capture such errors more accurately with articulographic or acoustic measures and analyse them statistically in order to get a more precise picture of which clusters are the most stable and which ones the least.

Secondly, errors were only analysed on the level of the cluster as a whole. A more fine-grained analysis that investigates which of the two consonants of a cluster the error occurs on could shed further light on which consonants in a cluster are affected by phonological speech errors the most and why. (To a certain degree, the confusion matrix in Appendix B provides some insight.) It would also be insightful to conduct a similar study that includes more German consonant clusters and see if the results from the present experiment also apply to them.

It should also be kept in mind that it cannot be assumed with certainty that all productions in this experiment that deviate from the target are true slips of the tongue. In some cases, they might be due to memory deficiencies leading to the wrong target representation for the participant. In an attempt to overcome this shortcoming, subjects were instructed to provide comment when they were unsure if they correctly remembered the stimulus, and these productions were excluded from analysis. Nevertheless, in some cases, subjects might simply not have been aware of the memory error.

Since Stemberger (1991, 2004) argues that frequency effects (albeit on the phoneme level) become apparent in nonce words, while anti-frequency effects arise in real words, it would be interesting to compare the results from this study with data from an analogous study with real words. If Stemberger is right, the pattern of results in such a parallel study should deviate considerably from that found here.

Finally, it is surprising that no effect of cluster neighbourhood could be found, both in light of previous studies that showed the effects of lexical neighbours in speech production and the findings concerning cluster neighbourhoods from the perception experiments reported in Chapters 5 and 6.[30] One reason for the absence of a neighbourhood effect in the data from the tongue twister experiment might be that the measure of neighbourhood was too coarse. A distinction between friends and enemies, as made by Stemberger (2004; cf. Section 8.2.3), might reveal antagonistic effects of the two kinds of neighbours, which level each other out when occurring together.

[30] Note, however, that generalised frequencies instead had an inhibitory effect, thus assuming the role of competition that neighbourhoods usually have in comparable experiments.

9. General discussion

This dissertation aimed to answer two main research questions: on the one hand, to what extent the automatisms of speech processing are based on universal principles, like the SSP, and, on the other hand, to what extent they are based on experience with a specific language and its frequency distributions. For this purpose, the production and perception of German initial consonant clusters in the context of pseudowords was investigated in separate experiments. For perception, an identification-in-noise paradigm was used. The same experiment was then conducted with a group of native and a group of non-native listeners. For production, a tongue twister paradigm was employed.

The results of the three experiments and their implications for a) the importance of different frequency measures and b) our knowledge of phonological representations will be discussed in this chapter. Error rates for the 16 consonant clusters used in all three experiments will be compared across experiments and the similarities and differences interpreted. Furthermore, the main results of the individual experiments will be summarised and conclusions drawn with respect to their generalisability over modalities and language background.

9.1. Comparison of perception and production results

In the L1 perception experiment, (log-transformed) cluster frequency had a strong facilitative effect on error patterns. Its model estimate was the highest except for that of sonority violation, which was not interpretable in a straightforward manner: SSP violations led to lower error rates, contrary to expectations. The situation was very similar in the L2 perception experiment. Here, too, the effect of SSP violation was the opposite of the theory-based prediction, while German cluster frequency had the predicted facilitative effect. However, there was an interaction between German and English cluster frequencies, which indicates that the facilitative effect of German frequency is very strong for clusters that are uncommon in English but not for those that are common in English. For clusters that are common in English, there was even a slight trend towards a detrimental effect of German frequencies on error rates. The frequency of a cluster is, therefore, a very good predictor of perception accuracy in both experiments, although the issue is more complicated in L2 perception due to the interaction with L1 frequencies (which themselves showed no significant effect).

Contrarily, the size of the frequency effect was rather small compared to that of the other variables in the production experiment. Here, direct competition between speech plans (as present in the metathesis stimuli) and immediate articulatory factors, like the total number of consonants without intervening vowels (captured by the predictor coda in previous syllable), were far more influential than the more abstract factors, frequency and sonority. In contrast to the perception experiments, SSP violation did not yield a significant effect, but the trend was in the same direction: clusters that violate the SSP showed lower error rates. The finer measure of sonority distance between consonants was significant at one level, namely a sonority distance of 1, which includes the notorious sibilant–stop clusters and the /ʃ/ + nasal clusters. Both the universal and language-specific sequence preference bias thus had a stronger influence on perception than on production (although the sonority effect was paradoxical). One possible explanation for the stronger frequency effect in perception is that there is greater room for conscious decisions than in the production experiment. When in doubt about what they heard, participants might have made (semi-)conscious decisions to go with the high-frequency (HF) clusters. The production study, on the other hand, tapped into automatic processes and left less room for conscious or semi-conscious biases. It is known that both sonority and language-specific frequencies of phonotactic sequences play a major role in judgement data, the task most heavily influenced by conscious decisions (Albright, 2009; Coleman & Pierrehumbert, 1997; Jarosz, 2010; van de Vijver & Baer-Henney, 2012). It seems that their influence diminishes the more automatised the task becomes. In the perception task, which is less conscious than judgement, the influence is still reasonably high, although arguably less so than in rating tasks. The speeded production task, in contrast, did not contain an element of uncertainty as a doorway for such biases. Since all errors are based on unconscious slips due to misparses of the phonological plan or articulatory difficulties under time pressure, the effects of cluster frequency were much smaller and the effects of SSP violation absent altogether.

These findings can be interpreted as follows: the frequency effect has two components, an unconscious element which might be based on representational strength or connection weights and a (semi-conscious) bias towards HF structures in situations of uncertainty. In the perception experiment, both components probably added to produce a larger frequency effect. In the production experiment, only the purely unconscious component showed an effect. Since the production targets were known to the subjects, there would have been no advantage in aiming for a HF pattern—contrary to the situation in perception, whereby opting for a HF cluster increases the chances of correct identification. If only one aspect of frequency effects was active in the tongue twister task but two in the identification task, this can explain why the frequency effect was much stronger in perception.

Concerning sonority sequencing, the reversed effect (i.e., the inhibitory effect of SSP conformity) in all experiments indicates that sonority as a sequencing principle does not take on a facilitative role in many processing tasks. Contrary to its role in certain kinds of impaired speech production, L1 acquisition, and perception of illegal sequences, the SSP does not seem to guide unimpaired and fully developed speech processing. The reversal of the predicted effect was traced back to sibilant–stop clusters as stronger (i.e., as behaving more unit-like) than stop–sibilant clusters in both modalities. Similar observations have been made for perception by Davidson and Shaw (2012) and Bond (1971). Future research should investigate the origin of this sonority hierarchy reversal further. For perception, it is likely connected to the high acoustic prominence of sibilants, which seems to be more relevant to speech modulation than sonority with its lack of clear physical correlates (Baroni, 2013; Henke et al., 2012). For production, far more research is needed to investigate why there is such a strong discrepancy between the two types of clusters in terms of production difficulty. If these problematic clusters are left out of the analysis, however, a slight effect of sonority sequencing emerged in production: in reduction errors, consonants were deleted so as to maximise syllable-initial sonority rises. It should be investigated whether the same is true for perception. Since SSP violations in many languages are limited to sibilant–stop clusters, the finer measure of sonority distance would be an adequate basis for this investigation.

Having compared the general strength and direction of the frequency and sonority effects across experiments, the error rates for the individual clusters in production and perception will now be inspected in more detail. If these are similar, it would indicate that frequency and/or sonority are influential in speech processing at a higher level because phonemes that are easy to perceive are not necessarily easy to produce and vice versa. For example, [s] is one of the most difficult phonemes in terms of articulation (Baroni, 2014; Tilsen, 2016) but is easily perceptible and noise-resistant due to its high intensity, especially at high-frequency bands (Reetz & Jongman, 2009; Wright, 2001).

Equally good processability in production and perception would therefore most likely be related to a cluster's mental representation.

Only L1 perception will be considered here because the production experiment included only L1 speakers. Figure 9.1 shows a comparison of the error rates.

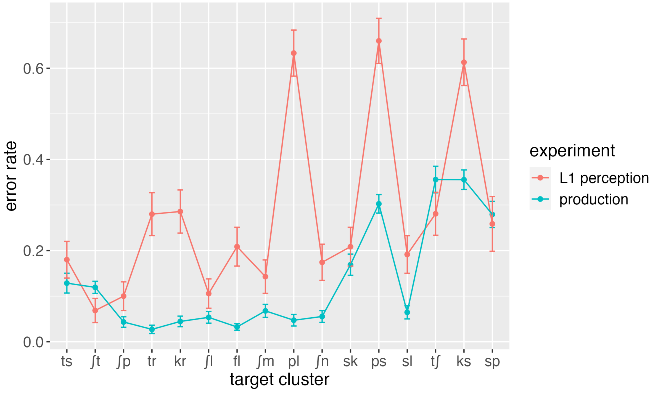


Figure 9.1.: Error rates over target clusters in L1 perception and production (Clusters are arranged according to their frequencies from most to least frequent.)

As can be seen, error rates were much higher overall in perception than in production. In order to have a better idea of where they diverge in the two experiments, the values were adjusted. This was done by raising the error rates for produced clusters and lowering the error rates for perceived clusters by 0.0682568 (which is half the difference between the overall error rates in the experiments; overall error rate production: 0.138248, overall error rate perception: 0.2747616). The result can be seen in Figure 9.2.

For both production and perception, there is a tendency for higher error rates towards the right (i.e., low-frequency [LF]) end of the cluster range, which reflects the frequency effect found in both experiments. However, in both curves, there are also deviations from this

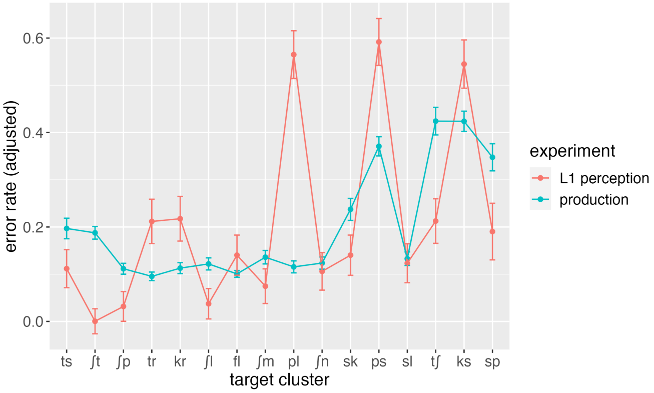


Figure 9.2.: Adjusted error levels over consonant clusters in L1 perception and production
(Clusters are arranged according to their frequencies from most to least frequent.)

trend. Moreover, it can be seen that variance is higher in the perception than in the production data. In the production experiment, the error rates for the high-to-mid frequency range are very similar, while they vary considerably in the perception experiment. The exceptions in the production experiment are the two onsets with the highest frequency, /ts/ and /ʃt/, as having unexpectedly high error rates, while /sl/ and /sp/ have unexpectedly low error rates when considering their frequency. The most plausible explanation for the high error rates of /ts/ and /ʃt/ is not that they are difficult to articulate, but that this is an artefact of the specific cluster pairing in the experiment (cf. Table 8.1 on p. 265 and Figure 8.1b on p. 276). The comparison of error rates for /ʃt/ in /ʃt–tʃ/ vs. in /ʃt–ʃn/, in particular, supports this interpretation. For /ts/, only data from one pairing is available, unfortunately—namely, its pairing with /ks/. As /ks/ was one of the most error-prone clusters in the experiment, however, it is reasonable to assume that the difficulty spilled over to its partner cluster /ts/ in the form of exchange errors, even though it is not a metathesis pair.

The two clusters with disproportionately low error rates, /sl/ and /sp/, are also exceptions (troughs) in the perception experiment. This suggests that there may be something about their mental representations that makes them relatively error-resistant (the alternative being pure chance). The error rate of /sp/ would probably have been even lower had it not been paired with /ps/. The two clusters have in common the fact that the /s/ is in C1 position. This situation is reminiscent of the special status of sC clusters in perception. However, it is difficult to determine with certainty if sC clusters are also more error-resistant in the production experiment because most of the clusters have very low error rates anyway, which produces a kind of ceiling effect, and the error rates of /ʃt/ and /sk/ are increased due to their pairing in the stimuli. Nevertheless, it can be speculated that sC clusters generally have better representations, irrespective of modality. /sl/ may also have profited from the relatively high frequency of the neighbouring (i.e., phonologically and phonetically similar) cluster /ʃl/. This would mean that the frequency effects are at least partly grounded in feature-based generalisations. The issue will be taken up again in Section 9.2.2 below.

In the perception experiment, there are more peaks in the error curve than in the production experiment, of which the most striking is /pl/. While this cluster has an error rate of 5% in the production experiment, its *perceptibility* is about as bad as that of the notoriously difficult clusters /ks/ and /ps/.[1] It is safe to assume that this is due to auditory reasons. As explained in Chapter 4, stops heavily rely on (segment-)external auditory cues for their recognition—mostly formant transitions, which are most easily detectable if the following phoneme is a vowel. Therefore, stop consonants followed by other consonants have a

[1] These two have proven to be weak in all experiments. They were also frequently misperceived in the production experiment—which did not contain any noise during stimulus presentation—as evident from wrong productions in the familiarisation phase. These trials were excluded from analysis, of course.

perceptual disadvantage. This is also evident from the "outlier" status of /tr/ and /kr/. Surprisingly, /tʃ/ fits into the general frequency-based pattern: its error rate is in the expected range on the basis of its frequency, even though /ʃ/ is a worse carrier of external cues than the liquids in the other stop-initial clusters, and /tʃ/ is part of a metathesis pair. Why it is so much less error-prone than the other stop-initial clusters remains unclear; its error rate in the production experiment was appropriate for the cluster's frequency.

Summing up, the results from perception and production share some aspects in common. On the whole, HF clusters are processed more accurately. Specifically, both experiments show striking similarities with respect to the very high error rates in the class of non-native stop–sibilant clusters. This suggests that it might be something about the class as a whole that led to the high error rates. The frequency effect was stronger in the perception experiment than in the production experiment. The same is true for the sonority effect, which was not even significant at all in the production experiment. In both modalities, however, there is an *inhibitory* effect of SSP conformity, the opposite of what would be expected. Sonority theory, therefore, makes no meaningful contribution towards explaining the data.

Apart from these commonalities between frequency and sonority, there were also patterns in the error data that can be attributed to the specific tasks at hand. In both perception experiments, the (psycho-)acoustic disadvantage of an initial stop followed by another consonant was very clear in the data, and the intensity of an onset was a significant predictor for its recognition. In production, it was not only ease of articulation (operationalised as increased cluster length due to a coda consonant in the previous syllable) that characterised the error patterns beyond phonotactic effects, but primarily the combination of clusters that co-occur in a stimulus pair: the effect of consonant metathesis showed the strongest influence. This means that the greatest difficulties for the participants arose at the planning stage, in which the order of consonants was confused, not at the stage of articulation—although the complexity effect suggests that the latter was also prone to disturbances. The significant interaction between frequency and metathesis indicates that the frequency effect, which was significant only for non-metathesis pairs, also arose at this level. The fact that only type frequencies showed an effect serves as further evidence for this conclusion.

9.2. Which frequency measures are relevant?

9.2.1. Type or token frequencies

The question of which frequency measure is relevant for various tasks in speech processing has been intensively debated (for an overview, see Hofmann et al., 2007). In the present studies, the main focus was on type frequencies since they have been found to be most closely connected to phonotactics (Hay et al., 2004). However, since the issue of type vs. token frequencies is still far from settled in psycholinguistics, the role of token frequencies was tested in all three experiments, too. This was done by replacing type frequencies with token frequencies after model fitting had been completed. In the production experiment, only type frequencies (both measures, i.e., taken from CELEX and elexiko) showed a significant effect on production accuracy, while token frequencies did not. In the L1 and L2 perception experiments, in contrast, both type (CELEX and elexiko) and token frequencies (CELEX and CLEARPOND) yielded significant effects, but those based on token frequencies were smaller and more variable. This suggests that type frequencies are the more relevant measure both in the production task and the perception task used here. It is conceivable that the effect of token frequencies is an artefact of the high correlation between the two frequency measures and thus failed to surface in the already weaker frequency effect in production.

Nonetheless, it is worth considering whether the token frequency effect in the two perception experiments is a genuine, independent effect. In visual word recognition, there is a strong dissociation between type and token frequency effects: while type syllable frequency has a facilitating effect, the effect of token syllable frequency is inhibitory in lexical decision (Conrad et al., 2008). The facilitating effect of type frequency is in line with the present results on perception. However, there was no inhibiting effect of token frequencies, but rather a facilitating effect that was weaker and more variable than that of type frequency. Conrad et al. (2008) ascribed the facilitating effect of type frequencies to the prelexical processing stage and the inhibitory effect of token frequency to the lexical stage. The results of their study are therefore not in direct conflict with the results of the present studies, which do not include a lexical processing component.[2] Specifically, Conrad et al. (2008; p. 320) assumed that the facilitating effect of type frequency is due to a "prelexical processing advantage for [...] units of high typicality". It is probably also this typicality, which is associated with type frequencies that facilitated recognition of HF consonant clusters here. It is unlikely that token frequencies, which have an inhibiting effect on the lexical level, would show a facilitating effect independent of type frequencies on the sublexical level.

When it comes to speech production studies, many only test for effects of token frequency (e.g., Bose et al., 2007; Pouplier, Marin, Hoole, et al., 2017; Riecker et al., 2008; Tremblay et al., 2016). In light of the present results, it must be questioned whether token frequencies re-

[2] However, they ultimately related both effects to the mean neighbourhood frequency. In the present experiments, the effect of neighbourhood is independent of that of target type or token frequencies. Both neighbourhood frequency and cluster frequency yielded significant effects in the regression models. Nonetheless, the idea—also found in (Vitevitch & Luce, 1999)—that frequency-based facilitation takes place at the prelexical level and inhibition at the lexical is plausible and can account for the facilitating effect of cluster frequencies in the perception experiments reported here. Nonetheless, the question of whether it is type or token frequencies that are responsible for this effect is not resolved.

ally are the most suitable frequency measure in all cases. The results of the production experiment reported here suggest that, at least for pseudowords, type frequencies, not token frequencies, predict production accuracy. It seems to be a consonant cluster's typicality in a language that determines its representational strength and consequently the ease with which it is produced. According to connectionist models (e.g., Dell's Spreading Activation Theory), this effect is due to these clusters being connected to many different syllables and lexemes and hence receiving activation from many nodes. In a study of past-tense formation, type frequencies also outperformed token frequencies for pseudowords, while overall, token frequencies modelled human behaviour best (Del Prado Martín et al., 2004). This underlines the varying roles of type and token frequencies for real words vs. pseudowords in different processing domains.

In addition to the type and token frequencies of sublexical units, transitional probabilities (TPs) between phonemes are regularly employed in psycholinguistic studies. The present study adds to the evidence (cf. Kawamoto & Kello, 1999) that, if TPs are chosen as a measure of probability, one should take into account backward TPs in addition to forward TPs. The results of the present experiments showed that the consonant in C2 position can bias both perception and production of the consonant in C1 position towards a HF transition. This was shown for /ks/ > /ts/ errors and, to a lesser degree, for /sl/ > /ʃl/ errors.

9.2.2. Segment-specific or generalised frequencies

Albright (2009) found both segment-based and feature-based phonotactic probability to have an effect on acceptability ratings. He concluded that there are multiple levels of evaluation. Likewise, in speech segmentation and lexical acquisition, feature-based abstractions have been found to play a role alongside segment-specific probabilities (Boll-Avetisyan, 2012). In the present studies, the generalised (i.e., feature-based) frequencies did not contribute significantly to explaining the data. One obvious difference between the studies concerns the tasks used: while Albright's subjects were required to give acceptability ratings of the clusters (a meta-linguistic task), the subjects in the present studies had to produce and identify the clusters. However, there is no plausible explanation as to why generalised frequencies should play a larger role in a meta-linguistic task than in tasks that tap more into on-line processing and direct access of representations.

A more likely explanation lies in the type of clusters used. Feature-based generalisations are more effective for unattested clusters (Albright, 2007b). When a phoneme sequence is known to the language user, they can rely on its specific distribution in the language. Generalised frequencies would only distort this optimally specific frequency knowledge. It is for unknown sequences that feature-based generalisations are the most beneficial. This probably holds for processing advantages as well and could explain why there was no overall effect of generalised cluster frequency, but /sl/ (a non-native, although attested, cluster) seems to have profited from the high frequency of the auditorily and articulatorily similar cluster /ʃl/ in all experiments. To a lesser degree, this might have even been the case for the marginal and non-native clusters /sk/ and /sp/. It can therefore be concluded that, for attested, native clusters, it is segment-specific frequencies that influence ease of processing. For very marginal, non-native clusters, however, generalised frequencies based on natural classes might influence processing. This explanation can account for the differences in error rates between /sl/, and maybe /sk/ and /sp/, on the one hand, and the stop–sibilant clusters on the other.

9.2.3. Source corpus

This thesis also adds to the discussion around the adequacy of CELEX frequencies for psycholinguistic studies. It was shown that CELEX type frequencies are able to predict processing accuracy both in perception and production. As discussed in Section 8.6.1, CELEX lacks a number of relatively recent English loan words and therefore does not map frequencies of, for instance, initial /sl/ faithfully in terms of modern German usage. However, its frequencies are still accurate enough to model frequency effects in speech processing. Crucially, consonant cluster frequencies based on other sources—namely the CLEARPOND subtitle corpus and elexiko—did not achieve better model fits than CELEX frequencies.

In any case, frequency tallies used in psycholinguistic experiments can only ever be approximations. Just as people have different internal grammars (Dąbrowska, 2012), they probably also have different representations of the clusters used in the studies presented here. For example, somebody who knows a person called *Xaver* will have a much stronger representation of initial /ks/ than most other people. Likewise, a German psychologist, who hears and uses words beginning with /psyːço-/ very frequently, will have a stronger representation of initial /ps/, and so on. Therefore, the frequencies used as a basis for the investigations do not accurately reflect individual participants' internal frequencies. With respect to the present experiments, the emergence of a frequency effect in all three suggests that the frequencies used here serve as a good average of participants' individual frequencies.

9.3. Phonological representations

The nature of phonological representations is a matter of ongoing debate. Although no definite conclusions concerning their nature can be drawn on the basis of the three experiments, the insights that can be gained from a comparison of the results of the three experiments will be briefly discussed in this section. The issues addressed are 1) whether consonant clusters are represented as units, that is, whether they have their own phonological representation, 2) whether phonological representations are shared between production and comprehension, and 3) the extent of separation between L1 and L2 sublexical representations, namely, whether L2 representations are influenced by their L1 counterparts.

The results of the three experiments support the notion that there are representations of consonant clusters that behave as holistic units. The error patterns for the onset clusters cannot be explained by their phonological make-up (e.g., the frequencies of their component phonemes) alone. Moreover, the variance in error rates within a cluster was much smaller than the variance between clusters (see Figure 9.1). In contrast to syllable frequencies, cluster frequencies had a significant influence on processing accuracy in all three experiments. However, the effect of phoneme frequency (i.e., the summed frequency of the two onset phonemes) was even stronger than that of cluster frequency in the production experiment. This means that, even though cluster frequencies play a role in processing and consonant clusters seem to be processed holistically at least in some cases, the primary sublexical unit would appear to be the phoneme. The consonant cluster is most likely a unit that has a mental representation in addition to that of the individual phonemes, which is used in cases where it is more efficient to refer to larger units, for example, when activating HF clusters as single units. This supports the notion that several sublexical units are used alongside each other during speech processing (e.g., Shattuck-Hufnagel & Klatt, 1979). One way in which this coexistence might work is modelled in ART, in which list chunks of various sizes exist. Importantly, larger chunks mask smaller ones in ART—something that seems to be at odds with the present results.

The concept of (onset) consonant clusters as representational units is also promoted by Bond (1971), Cutler et al. (1987), and MacKay (1972). Another possibility that receives support from the present data is that only some clusters are stored and processed as holistic units, while others are not (Berg, 1989). According to Berg (1989), a cluster's cohesiveness, and hence its unit-like behaviour, depends on the sonority values of the phonemes involved and the syllable structure. In the analyses in this dissertation, in contrast, it was shown that sonority relations between consonants are partly at odds with the degree of cohesiveness they displayed in the experiments. Sibilant–stop clusters behaved far more unit-like than stop–sibilant clusters and are more likely to have their own representation. At best, the hierarchy of consonant classes would have to be revised in order to account for the present results. The classification by Tzakosta (2009) seems to be a good starting point.

A more promising explanation for why some of the test clusters behaved more unit-like than others is provided by usage-based theory. It is likely that combinations of phonemes that are used together more often (i.e., clusters of a high frequency) are chunked together to a larger degree and behave more unit-like in processing. This means clusters undergo different degrees of entrenchment as a function of their frequency of use. In that case, HF clusters would be treated as units, while LF clusters are merely sequences of phoneme units. All in all, the present results support the usage-based view that phonological representations of different sizes coexist and are employed during speech processing according to the situational needs.

The question of whether phonological representations are shared between production and comprehension is still controversial. Positions for shared (Liberman & Mattingly, 1985; MacKay, 1982) and separate (e.g., Klatt, 1981; Warker et al., 2009) representations are both backed up by empirical research. Note, however, that most proponents of separate representations assume some kind of link between input and output systems, with the possibility of one (partially) influencing the other so that representations in one system are strengthened by the other (Kittredge & Dell, 2016; Warker et al., 2009; Zamuner et al., 2016). Based on their experiments, which investigated the transfer of phonotactic learning from the speech perception system to the speech production system, Warker et al. (2009) assumed that there are separate phonological input and output representations. The results obtained in the experiments presented in this dissertation suggest that if representations are separate—a question that this dissertation did not seek to answer and that cannot be resolved using the methods applied here—then the principles that underlie them must be substantially similar, which is shown by their largely parallel effects in the perception and production experiments. Note that this is not in conflict with Warker et al.'s findings; rather, it is a possibility they acknowledged, too. Not only is there a processing advantage for HF clusters in both experiments, but there are also striking similarities in the exceptions to this pattern, namely the low error rate of /sl/ and the exceptional status of stop–sibilant clusters. Furthermore, the unexpected anti-sonority effect—SSP-violating clusters as entailing a processing advantage—surfaced both in production and in perception: in production, this trend did not reach significance, whereas it was found to have a significant effect in perception. All of this can be taken as an indication of shared phonological cluster representations used during both tasks; yet, it cannot be taken as clear evidence.

With respect to the question of shared representations in L1 and L2, the evidence from the perception experiments is mixed. On the one hand, the L2 listeners were very strongly influenced by German cluster frequencies (even more so than the native listeners), while at the same time the English frequencies did not show a main effect. This suggests a strong separation between L1 and L2 representations. The fact that experience with Southern German dialects did not facilitate the perception of /ks/ for the native listeners, and knowledge of Greek did not facilitate perception of /ps/ and /ks/, supports the interpretation of separate representations for different languages and varieties. On the other hand, the interaction between L1 and L2 frequencies shows that they are not completely independent. The effect of German frequencies was strongest for clusters that have a very low frequency in English or are not attested at all. For English-HF clusters, the German frequencies did not play a role. This might be an indication that the acquisition of German clusters and their distribution in the L2 is only possible through the lens of the L1, and thus they are only unaffected by L1 representations when there are none or when they are weak.

Moreover, L2 listeners also showed phonetic influences of their L1. For example, they made more voicing errors than the L1 listeners and reported /tʃ/ as /tr/ because, in their L1, the two sequences are phonetically similar. Again, this indicates that L2 clusters are learned through the lens of the L1, which has also been suggested by Lentz and Kager (2015). In this respect, it parallels the situation in phoneme category acquisition, which proceeds through the lens of the L1 (Best & Tyler, 2007; Escudero & Boersma, 2004; Trubetzkoy, 1939). Therefore, it is not possible to determine with certainty whether the English and German cluster representations of L2 learners are shared or separate. It is likely that the representations are separate but linked and can therefore influence each other.

9.4. Issues in the interpretation of the sonority sequencing effect

There are a number of things that should be kept in mind when interpreting the effects of sonority sequencing in the three experiments. The choice of the specific clusters used in the experiments may have contributed significantly to the results. A comparison of stop–nasal vs. stop–liquid clusters might have yielded quite different results concerning sonority than the comparison of stop–sibilant vs. sibilant–stop clusters did. The contrast chosen here revealed an important limitation to the validity of sonority sequencing and provided potential explanations for it.

Recall also that the special status of sC clusters has been acknowledged by most researchers, and different proposals have been made in an attempt to demonstrate that they do not present true violations of sonority sequencing. However, if the sibilant is considered extrasyllabic, this should impede processing just as much as a sonority violation. Only if initial sC sequences were treated as a single phoneme would their processing be expected to be as good or better as that of clusters that conform to sonority sequencing.

Albright (2007b) found that neither statistically learned sequencing distributions (of the target language) nor prior universal preferences alone can explain his rating data; but adding sonority-based universal preferences to language-specific learned preferences significantly improved the model. In the experiments reported in this thesis, in contrast, sonority-based sequencing biases did not add to the exposure-based explanation of error patterns. This discrepancy in results might be due to task differences or the fact that only attested clusters were used in the present study, whereas Albright (2007b) used both attested and unattested clusters. It is very likely that the effect of sonority sequencing is strongest in unattested clusters, mirroring a universal bias. Attested clusters, on the other hand, can be assumed to be overlearned, especially HF clusters, and experience in a specific language can overwrite any universal biases. In the present experiments, this effect became visible in the HF clusters /ʃt/ and /ʃp/, which had very low error rates in spite of their violating the SSP. The error rates for the mid-frequency (MF) cluster /sk/ and, in particular, the LF cluster /sp/ are considerably higher both in production and in perception. (It should be noted, though, that the perception data on the latter are not as reliable because of potential spelling confusions.) It can therefore be concluded that a possible universal bias against such SSP-violating clusters has yielded to distribution-based biases. However, if no direct distribution information is available to the language user, as is the case for unattested clusters, sonority biases might be able to exert an influence, as can be seen in Albright's data. In such cases, sonority-based preferences and preferences based on indirect distributional information from clusters that are featurally similar to the cluster in question seem to collaborate. In the data reported here, there are hints that preferences derived from feature-based generalisations do play a role for marginal, non-native clusters. Sonority-based preferences, in contrast, are not found, even for these non-native clusters.

It should also be kept in mind that the domain of sonority is core syllabification (Clements, 1990). Clements (1990) notes that exceptions to sonority principles occur in many languages and ascribes this to morphologically complex syllables. It could be argued that the language users tested in the experiments reported here did not show any sonority sequencing effects because they are familiar with the exceptions originating from a level after core syllabification. However, in languages like German, these exceptions apply primarily to the syllable coda (due to suffixes, as in *des Schrank+s* "of the cabinet"), and, since the clusters in the experiment were all onset clusters, such post-core syllabification processes should not have played a major role.

10. Conclusion

Speech production and perception are complex processes, yet, language users usually accomplish both tasks effortlessly and efficiently in everyday communication. In order to achieve this efficiency, the human speech processor exploits regularities on various linguistic levels that help in automatising subprocesses. This dissertation investigated mechanisms that might contribute to the automatisation of sublexical speech processing. Specifically, two sources of consonant sequencing preferences, which potentially facilitate the perception and production of initial consonant clusters, were contrasted and their relative influences investigated.

On the one hand, the frequencies of use of the consonant clusters might guide their processing: high-frequency (HF) clusters might be more automatised in production, while perception might be biased towards recognising them. According to usage-based linguistics, our experience with language shapes our mental representations of it and these altered representations, in turn, influence subsequent processing. For example, the representations of HF structures are strengthened by their repeated use and are easier to process in future encounters. Elements that are commonly used together undergo further entrenchment; this means that they form a larger unit, which can be processed in a more automated way without attention paid to the individual components. While this mechanism has been investigated extensively with respect to multi-word sequences, far less attention has been given to the sublexical level. It was argued here that (common) consonant clusters are similarly chunked together and therefore represent a unit in speech processing, Consequently, HF clusters should have stronger representations and should be processed more easily; this means they are less susceptible to errors than low-frequency (LF) clusters and, at the same time, are more often the outcome of an error (i.e., produced or perceived instead of the target cluster).

On the other hand, universal biases for phoneme sequencing have been found, which provide a processing advantage for universally preferred structures, at least under certain circumstances. One prominent example is the Sonority Sequencing Principle (SSP), which states that the phonemes of a syllable should rise in sonority towards the syllable nucleus and decrease from then on. Sonority is a concept from phonological theory that is broadly correlated with a phoneme's intensity or the aperture of the oral cavity during production. A more gradual distinction is made by the Sonority Dispersion Principle (SDP), according to which sonority should rise maximally in the syllable onset and fall minimally at the end of the syllable. Hence, among syllable-initial consonant clusters that conform to the SSP, those with a larger sonority distance between consonants are more well-formed than those with a smaller sonority distance. Initial consonant clusters that are illegal in a given language have been found to be dispreferred by speakers of that language as a function of the degree of deviation from the SSP and the SDP. SSP-violating clusters are also prone to misperception and are subject to more errors in impaired speech production, especially in patients with apraxia of speech. This dissertation investigated whether SSP-conforming clusters have a processing advantage that also holds when all clusters are legal in the listeners' native language (L1)/target language and whether the advantage in production extends to healthy individuals. If this is the case, clusters that conform to the SSP should be produced and perceived more accurately than SSP-violating clusters, and they should act as repairs of SSP-violating clusters in errors.

In order to explore the relative influence of these two factors—cluster frequency and sonority sequencing—on sublexical speech processing, perception and production of the same 16 legal German clusters (/ts/, /ʃt/, /ʃp/, /tr/, /kr/, /ʃl/, /fl/, /ʃm/, /pl/, /ʃn/, /sk/, /ps/, /sl/, /tʃ/, /ks/, and /sp/) in syllable onset position was tested experimentally in the context of pseudowords. Even though /t͡s/ is considered to be an affricate in German, it was included in the set of test clusters to allow for a comparison with the structurally parallel sequences /ps/ and /ks/.

10.1. Summary of results

Chapter 5 examined the perception of initial consonant clusters by native listeners. Participants heard monosyllabic pseudowords starting with the test clusters embedded in multi-talker babble and were instructed to freely transcribe what they heard. In addition to a strong effect of intensity, listeners were clearly influenced by the frequencies of consonant clusters: the lower the frequency of a cluster, the higher the error probability; and the higher the frequency of a cluster, the more often it acted as an intruder in misperceptions. In contrast to the facilitating effect of the target cluster's frequency, the frequencies of neighbouring clusters showed an inhibitory effect: the higher the summed frequency of all neighbouring clusters, the more difficult it became to identify the target. This was interpreted as an indication of the dynamics of activation and competition in speech perception: HF clusters are more strongly activated and are therefore easier to perceive. Concurrently, neighbouring clusters add competition to the identification process. The more strongly activated the neighbouring clusters are as a result of their own frequency, the stronger the competition is; this inhibits identification of the target cluster and leads to higher error rates.

Contrary to expectation, perception was not hampered for clusters that violate the SSP. On the contrary, SSP-violating clusters were recognised correctly more often than SSP-conforming clusters. This was due to two reasons: firstly, sibilant–stop clusters—which violate the SSP since sibilants are higher in sonority than stops—had very low error rates. Secondly, stop-initial clusters had the highest error rates, even though voiceless stops with their minimal sonority value should be ideal in syllable-initial position. There is an acoustic explanation for the low perceptibility of stops in initial position: stop consonants have weak internal acoustic cues and hence rely on neighbouring segments which are good carriers of their external cues. Consonants, and especially sibilants, however, are bad carriers of such cues. As a result, recognition of a stop followed by another consonant, particularly a sibilant, is poor. It was therefore argued that the SSP is not, in itself, a tenable principle in native speech perception. Rather, its success in explaining some perceptual phenomena is due to the correlation of sonority, at least to a certain degree, with acoustic and perceptual factors. An account based on cue robustness would reverse the positions of sibilants and stops on the hierarchy, resulting in correct predictions concerning the perceptibility of the consonant clusters used in the experiment. In order to explain perceptual phenomena, a cue robustness account is therefore preferable to phonological principles based on abstractions, such as the SSP.

Chapter 6 addressed the question of whether these results also apply to second-language (L2) listening. This situation is potentially very different from L1 listening. L2 listeners, who are less familiar with target-language phonotactic distributions than L1 listeners, may show stronger universal biases. Moreover, they might be affected by phonotactic distributions in their native language in addition to the distributions found in the target language. Past research has shown that L2 listeners can be influenced both by their native language phonotactics and target language phonotactics in the listening process. In order to examine to what extent L1 frequencies, L2 frequencies, and sonority sequencing affect L2 listening and how they interact, Experiment 1 was repeated with a group of Australian learners of German;

German levels ranged from intermediate to advanced. Their behaviour was remarkably similar to that of the L1 listeners: they were strongly influenced both by acoustic factors and German cluster frequencies (target as well as neighbourhood). As a matter of fact, the effect of German cluster frequencies was stronger for the L2 listeners than for the German group. The frequencies of clusters in their L1, English, on the other hand, did not show a main effect. Nevertheless, L1 cluster frequencies modulated the effect of German cluster frequencies: German frequencies had the greatest influence on perception accuracy for clusters with a very low frequency in English, while clusters with a high frequency in English were recognised equally well, irrespective of their frequency in German. The fact that German frequencies yielded a main effect, while English frequencies did not, indicates that the sublexical frequency effect in speech perception—also found in Experiment 1—is based on target-language frequencies rather than L1 frequencies. The Australians were able to use their knowledge specifically about target-language (i.e., German) distributions to guide the listening process. However, they were not completely unaffected by their native phonotactics.

The stronger manifestation of the German frequency effect was explained by learners' skewed representations of German cluster frequencies. For example, even with the reduced input when compared to native speakers, L2 listeners are probably very familiar with the HF clusters /ʃt/ and /ʃp/, but might never have encountered the LF clusters /ks/ and /ps/. The perceptual illusions often found in L1-illegal structures were thus limited to clusters that are simultaneously illegal in the listeners' L1 and highly marked in the L2 in the present experiment. Furthermore, the clusters in question represent a cluster class (stop–sibilant) that is inadmissible in their L1 altogether. It was also suggested that L2 listeners rely on top-down information to a larger degree than L1 listeners because they are less experienced in parsing the language-specific acoustic cues for consonant identification. This might have further contributed to the enhanced effect of German frequencies.

With regard to sonority sequencing, the L2 listeners showed the same perceptual advantage for SSP-violating clusters as the L1 listeners, which can be attributed to the favourable cue recoverability of sibilant–stop clusters, particularly in comparison to their reversed counterparts. It was therefore concluded that sonority sequencing is no more influential in L2 listening than it is in L1 listening. However, this finding cannot be generalised to all universal phonological principles: L2 listeners did show an effect of Net Auditory Distance (NAD), a more fine-grained alternative to the SSP, which is grounded in Beats-and-Binding phonology, and has a strong foundation in perceptual and articulatory phenomena. The superiority of NAD over sonority was interpreted as an indication that in order for phonological principles to be relevant to psycholinguistic data, they need to refer to sufficiently fine-grained measures and be developed based on psycholinguistic processes.

Taken together, the two perception experiments demonstrated that L1 listeners and L2 listeners are to a large degree susceptible to the same influencing factors in sublexical processing, namely acoustic parameters and sublexical distributions in the target language. A notable difference between the groups is that L2 listeners, but not L1 listeners, are influenced by the NAD of sublexical sequences. This suggests that they are indeed affected by universal sequencing principles to a larger degree than L1 listeners—provided that these principles are based on psychoacoustically realistic measures. Moreover, for L2 listeners, the influence of target-language frequencies is modulated by L1 distributions.

Having established that cluster frequencies and neighbourhood frequencies play a role in perception, whereas sonority sequencing does not have a facilitating effect, Chapter 8 explored the question to what extent this is true for speech production as well. The 16 consonant clusters were arranged in pairs of phonologically similar clusters, which served as onsets for pairs of stimulus pseudowords (e.g., /ʃpɛl ʃmɛl/). Some of the cluster pairs were minimal pairs, meaning that only one phonological feature in one of the consonants differed (e.g., /ts–ks/), whereas a minority were so-called *metathesis pairs*, which consist of the same two consonants in a reversed order (e.g., /sk–ks/). During the experiment, participants repeated the auditorily presented stimuli four times at a fast rate (144 beats per minute), which was indicated by a metronome. This set-up caused a tongue twister effect.

As in the perception experiment, processing was facilitated for HF clusters, but the effect was smaller than in perception. The greatest influence on error rates was caused by the pairing of the stimuli: metathesis pairs were far more difficult to produce correctly at the required speed than all of the other stimulus pairs. This was interpreted as a sign of increased competition during the planning phase in cases of identical consonants that alternate in a sequence of syllables. Essentially, this means that it is not so much a consonant's or consonant cluster's inherent difficulty that determines production effort, but rather the context as a whole. Unlike impaired populations and children in the language acquisition phase, the healthy adult subjects in the tongue twister experiment did not benefit from syllables' adherence to sonority sequencing; neither sonority distance between consonants as a fine-grained measure nor SSP violation can account for the resulting error patterns. It was therefore concluded that, even under increased processing demands, as found in the tongue twister paradigm, healthy adults' prelexical speech planning is not guided by sonority sequencing. Instead, language-specific cluster frequencies take over this facilitating role. This was taken as an indication that overlearning a specific phonotactic system overrides universal sequencing biases.

As in perception, obstruent–obstruent clusters showed an interesting error pattern: sibilant–stop clusters were relatively error-resistant (while having very high rates of false positives), whereas stop–sibilant clusters had by far the highest error rates—the exact opposite of what sonority theory would predict. Since the characteristics that make a consonant difficult on a lower level differ inherently between production and perception, this similarity was ascribed to the mental representations of these clusters. It was concluded that sibilant–stop clusters have strong mental representations and behave in a more unit-like manner, while stop–sibilant clusters have weak or even lack mental representations as clusters; their components are thus likely processed separately.

10.2. General conclusions

Speech perception and production are very distinct processes rather than simply the inverse of one another. Yet, to some degree, they might be subject to the same mechanisms and principles. This dissertation explored two regularities that could potentially be exploited by language users to assist in making sublexical speech processing more efficient in both domains: language-specific sublexical (specifically, consonant cluster) frequencies and universal preferences for syllables that conform to the SSP.

The central question that this dissertation sought to answer is whether language-specific frequencies and/or sonority sequencing influence(s) perception and production of German consonant clusters. The main hypothesis was that consonant cluster frequencies have a significant influence on speech perception and production accuracy. In line with predictions from usage-based linguistics, cluster frequencies showed the predicted facilitating effect in all three experiments.

It was further hypothesised that clusters that violate the SSP would be harder to process in perception and production (although the influence of sonority sequencing was assumed to be smaller than that of language-internal frequency). Instead, SSP-violating clusters were found to be *easier* to process, significantly so in perception and as a trend in production. (This trend becomes significant only when sonority distance instead of SSP violation is used as a predictor.)

This indicates that, even when under great stress (as induced by the tongue-twister task and the identification task with an unnaturally high noise level), healthy adult speaker–hearers are not influenced by sonority-based universal preferences when processing legal (with respect to the target language) phoneme sequences. On the contrary, the clusters that conform to the SSP had higher error rates than those that violate it. Laeufer (1995; p. 260) notes that syllabification rules are determined by language-specific phonotactic rules in slow speech and by "the more relaxed sonority-based constraints alone" in fast speech. Based on this observation, one would also expect more sonority-improving speech errors in fast speech. That was not the case, nor was a resyllabification of the SSP-violating sequences visible in the data. The limited applicability of sonority principles to speech processing has been noted in the literature before and several more psychologically real measures have been proposed (Baroni, 2014; Dziubalska-Kołaczyk, 2014; Henke et al., 2012). One such measure, Net Auditory Distance (NAD; Dziubalska-Kołaczyk, 2014, 2019), was tested in the perception experiments. In contrast to sonority sequencing, it yielded significant results in line with the phonological theory behind it in L2 perception. It can therefore be considered superior to sonority theory when it comes to explaining speech perception phenomena. However, even NAD does not seem to play a role in L1 perception. In speech production, on the other hand, no alternative to the SSP could be tested. Even the most promising candidate, Ease of Articulation (EoA; Ziegler & Aichert, 2015), was too coarse to be applied in any meaningful way to the set of consonant clusters used here. Future studies will show whether a good replacement for sonority theory can be found in production, too.

Frequency, on the other hand, showed the predicted effect, albeit a weak one in production when compared to articulatory complexity (coda in previous syllable) and planning complexity (metathesis pairs). It was argued that the stronger manifestation of the frequency effect in perception is due to the task involving a stronger conscious component. Even though acoustic factors were most crucial to consonant cluster identification, it was demonstrated that listeners rely on top-down information, like sublexical frequencies, to a larger degree in noisy listening conditions than in quiet conditions.

The results from the present studies add to a long line of research showing that frequencies of different linguistic units influence their processing both in production and perception. As Pouplier, Marin, and Kochetov (2017; p. 472) state, "[s]peaking is probably one [of] the most complex but also one of the most overlearned behaviours in humans". Something similar can be said with respect to listening and processing perceived speech, an activity that happens effortlessly and cannot be suppressed. Humans are exposed to and use language throughout their lives, in most cases for many hours every day. Thereby, common linguistic structures become overlearned, and this leads to frequency effects of different linguistic units in various tasks. Moreover, it explains why prior biases, such as the ones based on sonority sequencing, usually do not have an effect on healthy adult speakers. The present dissertation demonstrated that consonant clusters, a sublexical unit which has received little scholarly attention thus far, show frequency effects comparable to those of other linguistic units. Language users, even second language users, are aware of these sublexical frequencies and are influenced by them during speech processing.

A. Perception experiments

A.1. Stimulus material

Table A.1.: Stimulus items in the perception experiments

CC	stimulus	status	block
ts	tsaʃ	test item	2
	tseːç	test item	4
	tsɛm	test item	5
	tsɪr	test item	3
	tsɔx	test item	1
	tsʊf	test item	2
	tsæːn	test item	1
	tsyːl	test item	5
	tsɔʏt	test item	3
	tsaɪ̯m	test item	4
ʃt	ʃtak	test item	1
	ʃteːm	test item	2
	ʃtɛf	test item	3
	ʃtiːn	test item	4
	ʃtɪŋ	test item	1
	ʃtoːt	test item	5
	ʃtɔr	test item	3
	ʃtuːx	test item	5
	ʃtœf	test item	2
	ʃtaʊ̯k	test item	4
ʃp	ʃpaːk	test item	4
	ʃpeːm	test item	3
	ʃpɛl	test item	2
	ʃpiːt	test item	5
	ʃpoːx	test item	5
	ʃpuːn	test item	2
	ʃpʊŋ	test item	1
	ʃpʏç	test item	3
	ʃpɔɪ̯s	test item	4
	ʃpaʊ̯f	test item	1
tr	traːs	test item	1
	tral	test item	4
	trɛt	test item	4
	triːs	test item	2
	trɪŋ	test item	5
	troːf	test item	3
	trɔʃ	test item	5
	truːl	test item	3
	træːp	test item	2
	tryːm	test item	1
kr	kraːx	test item	5
	krat	test item	3
	kreːs	test item	5
	krɛŋ	test item	1
	kriːl	test item	3
	kroːl	test item	2
	krɔf	test item	4
	kruːf	test item	1
	kræːk	test item	4
	kraʊ̯p	test item	2
ʃl	ʃlaːx	test item	3
	ʃlat	test item	5
	ʃleːç	test item	1
	ʃlɛr	test item	2
	ʃliːp	test item	1
	ʃlɪn	test item	4
	ʃloːm	test item	4

Table A.1 *(continued)*

continued on next column

continued on next column

Table A.1 *(continued)*

CC	stimulus	status	block
	ʃlʊf	test item	3
	ʃløːs	test item	5
	ʃlʏt	test item	2
fl	flaːk	test item	2
	flɛç	test item	5
	flɪs	test item	2
	floːt	test item	1
	fluːn	test item	3
	flʊŋ	test item	4
	flœp	test item	3
	flyːr	test item	4
	flɔ͜ɪk	test item	1
	fla͜ʊn	test item	5
ʃm	ʃmaːs	test item	3
	ʃmaŋ	test item	5
	ʃmɛl	test item	2
	ʃmiːn	test item	1
	ʃmoːt	test item	4
	ʃmɔx	test item	2
	ʃmuːf	test item	4
	ʃmøːl	test item	1
	ʃmɔ͜ɪç	test item	5
	ʃma͜ʊk	test item	3
pl	plaʃ	test item	1
	pleːs	test item	4
	plɛk	test item	3
	pliːt	test item	4
	plɪm	test item	2
	plʊt	test item	5
	plœn	test item	5

continued on next column

Table A.1 *(continued)*

CC	stimulus	status	block
	plyːl	test item	1
	plɔ͜ɪç	test item	2
	pla͜ɪs	test item	3
ʃn	ʃnat	test item	4
	ʃneːç	test item	1
	ʃnɛr	test item	4
	ʃniːl	test item	5
	ʃnɪf	test item	3
	ʃnoːx	test item	2
	ʃnuːk	test item	2
	ʃnœf	test item	5
	ʃnɔ͜ɪp	test item	3
	ʃna͜ɪm	test item	1
sk	skaŋ	test item	4
	skeːl	test item	3
	skɛf	test item	5
	skiːs	test item	3
	skɪr	test item	4
	skoːt	test item	2
	skɔr	test item	1
	skuːk	test item	1
	skæːs	test item	5
	skʏn	test item	2
ps	pasf	test item	1
	pseːt	test item	2
	psɛl	test item	2
	psiːr	test item	1
	psɪç	test item	5
	psoːx	test item	5
	psɔm	test item	3

continued on next column

Table A.1 *(continued)*

CC	stimulus	status	block
	psʊʃ	test item	3
	psøːt	test item	4
	psɔɪ̯n	test item	4
sl	slaːt	test item	1
	slan	test item	2
	sleːm	test item	5
	sliːʃ	test item	3
	slɪt	test item	2
	sloːn	test item	3
	slɔk	test item	4
	sluːp	test item	5
	slæːç	test item	1
	slyːf	test item	4
tʃ	tʃaːx	test item	4
	tʃaf	test item	2
	tʃɛŋ	test item	1
	tʃɪr	test item	1
	tʃoːt	test item	3
	tʃɔm	test item	2
	tʃuːp	test item	4
	tʃæːç	test item	5
	tʃɔɪ̯f	test item	5
	tʃaʊ̯s	test item	3
ks	ksan	test item	3
	ksɛp	test item	4
	ksiːt	test item	2
	ksɪm	test item	4
	ksoːf	test item	1
	ksɔp	test item	5
	ksuːt	test item	5

continued on next column

Table A.1 *(continued)*

CC	stimulus	status	block
	ksæːr	test item	3
	ksɔɪ̯l	test item	1
	ksaɪ̯n	test item	2
sp	spaːl	test item	2
	spat	test item	3
	speːç	test item	5
	spiːf	test item	5
	spɪk	test item	1
	spɔf	test item	3
	spuːx	test item	4
	spʊm	test item	2
	spaʊ̯t	test item	1
	spɔɪ̯n	test item	4
—	biːl	filler item	4
	bloːx	filler item	1
	blɪ̯n	filler item	5
	bɔl	filler item	5
	buːn	filler item	4
	brœf	filler item	2
	briːn	filler item	3
	dal	filler item	4
	dɪf	filler item	1
	dɔp	filler item	4
	duːl	filler item	2
	dryːt	filler item	3
	dreːf	filler item	4
	daʊ̯k	filler item	5
	gat	filler item	4
	giːf	filler item	1
	gloːn	filler item	5

continued on next column

Table A.1 *(continued)*

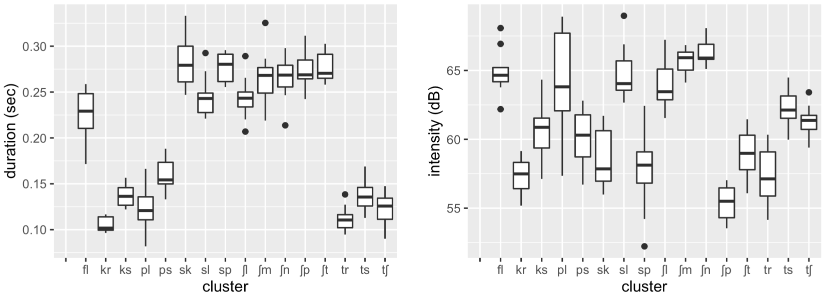
CC	stimulus	status	block
	gleːʃ	filler item	2
	grʊŋ	filler item	5
	græːʃ	filler item	4
	gɔɪ̯t	filler item	2
	gaʊ̯s	filler item	1
	peːk	filler item	3
	p͡fɛl	filler item	1
	pɪʃ	filler item	1
	prɔŋ	filler item	2
	prøːp	filler item	3
	p͡fɔɪ̯t	filler item	5
	tam	filler item	5
	teːʃ	filler item	3
	tiːl	filler item	2
	tʊp	filler item	1
	taʊ̯l	filler item	4
	taɪ̯f	filler item	5
	kleːʃ	filler item	5
	kɪʃ	filler item	2
	kluːf	filler item	3
	kvæːt	filler item	1
	kviːl	filler item	4
	knyːp	filler item	5
	knaɪ̯ç	filler item	3
	haːl	filler item	5
	han	filler item	1
	heːf	filler item	4
	hɛn	filler item	3
	hʊs	filler item	2
	haʊ̯k	filler item	2
	jaːt	filler item	1
	joːʃ	filler item	3
	jɔp	filler item	5
	jʊt	filler item	4
	jæːm	filler item	2
	fat	filler item	2
	fɪr	filler item	5
	froːp	filler item	4
	fɔx	filler item	4
	fruːn	filler item	1
	fʏs	filler item	3
	fɔɪ̯t	filler item	2
	vɛs	filler item	1
	vɪt	filler item	5
	voːf	filler item	2
	vʊŋ	filler item	3
	vʏç	filler item	2
	zeːf	filler item	1
	zɛm	filler item	5
	ziːʃ	filler item	4
	zɪl	filler item	3
	zoːx	filler item	1
	zɔt	filler item	2
	zuːf	filler item	3
	ʃrap	filler item	5
	ʃeːç	filler item	3
	ʃɛŋ	filler item	4
	ʃɪn	filler item	3
	ʃroːl	filler item	4
	ʃvuːs	filler item	2

continued on next column

Table A.1 *(continued)*

CC	stimulus	status	block
	ʃvøːt	filler item	1
	maːf	filler item	2
	meːk	filler item	5
	miːl	filler item	5
	moːt	filler item	3
	mʊt	filler item	4
	mɔɪ̯f	filler item	1
	neːç	filler item	4
	nɛf	filler item	5
	niːk	filler item	2
	noːx	filler item	1
	nɔʃ	filler item	2
	nʊp	filler item	3
	fraːn	filler item	3
	rak	filler item	5
	rɪç	filler item	4
	ruːp	filler item	1
	rʊt	filler item	1
	laːt	filler item	4
	laf	filler item	3
	loːx	filler item	2
	lɔp	filler item	1
	lʊt	filler item	3

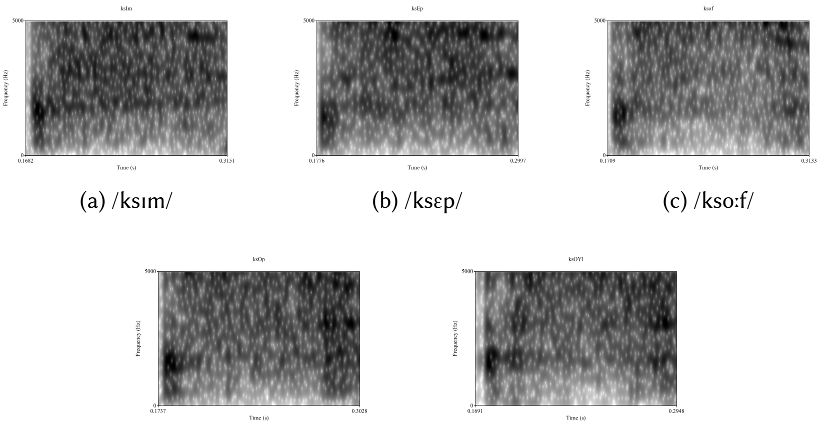
A.2. Stimulus characteristics



(a) Durations

(b) Intensities

Figure A.1.: Acoustic properties of stimulus onsets (test items only)



(a) /ksɪm/

(b) /ksɛp/

(c) /ksoːf/

(d) /ksɔp/

(e) /ksɔɪ̯l/

Figure A.2.: Spectrograms of the five /ks/ onsets that were misperceived by all L2 listeners

A.3. Questionnaires

A.3.1. L1 group

1. Alter:

2. Geschlecht: □ weiblich □ männlich

3. Händigkeit: □ RechtshänderIn □ LinkshänderIn

4. Studienfächer (auch frühere):

5. Geburtsort:

6. Sind Sie deutsche/r MuttersprachlerIn? □ nein □ ja

7. Sprechen Sie einen deutschen Dialekt/deutsche Dialekte?
   □ nein □ ja
   Falls ja, welche(n) und wie oft/gut? (siehe Skala)
   5 = (fast) täglich
   4 = ein- bis mehrmals pro Woche
   3 = selten, aber ich beherrsche den Dialekt aktiv
   2 = selten (Dialekt ist etwas eingerostet)
   1 = ich beherrsche den Dialekt kaum noch

8. Kommen oder kamen Sie regelmäßig mit einem oder mehreren deutschen Dialekten in Berührung (z.B. durch Familie, FreundInnen, PartnerIn)? □ nein □ ja
   Falls ja, mit welchem/n? Wie vertraut ist er/sind sie Ihnen?
   4 = sehr vertraut
   3 = relativ vertraut
   2 = nicht so vertraut
   1 = kaum vertraut

9. Sprechen Sie weitere Sprachen? □ nein □ ja
   Falls ja, welche und wie gut (siehe Skala)?

5 = nahezu muttersprachlich
4 = sehr gut
3 = gut
2 = mittelmäßig
1 = nicht besonders gut

10. Haben oder hatten Sie Hörprobleme? □ nein □ ja und zwar:

11. Liegen bei Ihnen in der Familie Hörschädigungen vor?
    □ nein □ ja und zwar:

Vielen Dank!

A.3.2. L2 group

1. Alter:

2. Geschlecht: □ weiblich □ männlich □ anderes

3. Studienfächer (auch frühere):

4. Was ist/sind Deine Muttersprache(n)?

5. Seit wie vielen Jahren lernst Du Deutsch?
   Seit ________________ Jahren

6. Wie alt warst Du, als Du anfingst, Deutsch zu lernen?
   ________________ Jahre alt

7. Wie schätzt Du Deine Deutschkenntnisse ein? (siehe separates Blatt) [official CEFR descriptions were provided on a separate sheet]

8. Warst Du schon einmal in Deutschland/Österreich/der Schweiz?
   □ nein □ ja → Falls ja, wann und für wie lange?

9. Wie oft übst Du folgende Tätigkeiten auf Deutsch aus? (pro Tag/Woche/Monat/Jahr)
    - lesen: _____ Mal pro _______________
    - hören: _____ Mal pro _______________
    - schreiben: _____ Mal pro _______________
    - sprechen: _____ Mal pro _______________

10. Kommst oder kamst Du regelmäßig mit einem oder mehreren deutschen Dialekten in Berührung (z.B. durch FreundInnen, PartnerIn, Medien)? □ nein □ ja
    Falls ja, mit welchem/n? Wie vertraut ist er/sind sie Dir?
    4 = sehr vertraut
    3 = relativ vertraut
    2 = nicht so vertraut
    1 = kaum vertraut

11. Sprichst Du weitere Sprachen? □ nein □ ja
    Falls ja, welche und wie gut (siehe Skala)?
    5 = nahezu muttersprachlich
    4 = sehr gut
    3 = gut
    2 = mittelmäßig
    1 = nicht besonders gut

12. Hast oder hattest Du Hörprobleme? □ nein □ ja und zwar:

13. Liegen bei Dir in der Familie Hörschädigungen vor?
    □ nein □ ja und zwar:

Vielen Dank!

A.4. Statistical models and additional effect plots

predictor	estimate	SE	z value	
(Intercept)	-1.088722	0.065481	-16.627	***
log cluster frequency	-0.656585	0.043648	-15.043	***
onset intensity	-0.044568	0.009098	-4.899	***
onset duration	-0.547733	0.043364	-12.631	***

Table A.2.: Output of acoustic model
Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$
Formula: `error~logFreqDE + ons.intensity + ons.dur + (ons.intensity + ons.dur | subjID)`

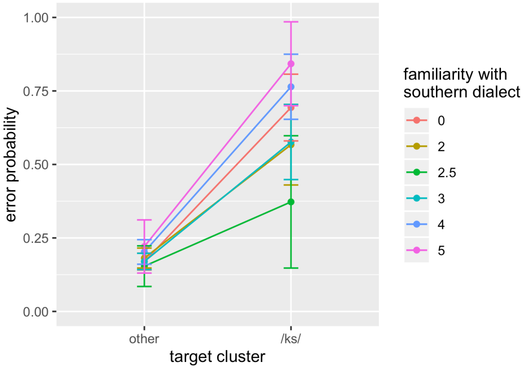


Figure A.3.: Effect of southern German dialect familiarity on identification of /ks/ vs. other targets

predictor	estimate	SE	z value	
(Intercept)	-1.589476	0.202746	-7.840	***
dialect familiarity 2	0.084115	0.253261	0.332	
2.5	-0.114052	0.541399	-0.211	
3	0.004179	0.224906	0.019	
4	0.218062	0.281263	0.775	
5	0.329043	0.536841	0.613	
target ks	2.406096	0.532542	4.518	***
dialect familiarity 2 × target ks	-0.631087	0.374468	-1.685	.
dialect familiarity 2.5 × target ks	-1.223685	0.787186	-1.555	
dialect familiarity 3 × target ks	-0.513178	0.334466	-1.534	
dialect familiarity 4 × target ks	0.141095	0.443133	0.318	
dialect familiarity 4 × target ks	0.530804	0.915994	0.579	

Table A.3.: Model predicting errors on /ks/ vs. other clusters from familiarity with southern German dialects
Significance codes: *** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1
Formula: `error~famdSouth * target.ks + (1 | subjID) + (1 | stimulus)`

predictor	estimate	SE	z value	
(Intercept)	-0.806832	0.355518	-2.269	*
log cluster frequency	-0.918065	0.261505	-3.511	***
generalised log cluster frequency	-0.006189	0.030565	-0.202	
SSP violation	-1.501473	0.619410	-2.424	*
summed neighbourhood frequency	0.196271	0.079175	2.479	*
onset intensity	-0.210692	0.048176	-4.373	***
salience-based wellformedness	-0.403770	0.504946	-0.800	

Table A.4.: Model output of the best-fitting model excluding data from southern dialect speakers
Significance codes: *** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1
Formula: `error~logFreq + son.vio + accNF + salience + logFreqGen + ons.intensity + (logFreq + son.vio + accNF + salience + ons.intensity|subjID) + (1|onset.targ/stimulus)`

predictor	estimate	SE	z value	
(Intercept)	2.15867	4.33456	0.498	
German log cluster frequency (type)	0.53915	0.71567	0.753	
English log cluster frequency (type)	-2.06565	3.70483	-0.558	
SSP violation (ref. level: no violation)	-1.53392	0.76252	-2.012	*
summed neighbourhood frequency	0.23821	0.08246	2.889	**
NAD difference	-0.34548	0.19979	-1.729	.
onset intensity	-0.17245	0.06040	-2.855	**

Table A.5.: Model summary of regression run on native English clusters only)
Significance codes: *** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1

predictor	estimate	SE	z value	
(Intercept)	-0.38337	0.27754	-1.381	
German log cluster frequency (type)	-2.08692	0.30676	-6.803	***
English legality	0.28887	0.35846	0.806	
SSP violation (ref. level: no violation)	-2.39632	0.46651	-5.137	***
summed neighbourhood frequency (German)	0.22552	0.05517	4.088	***
onset intensity	-0.16801	0.03547	-4.736	***
NAD	-0.40965	0.10605	-3.863	***
German log cluster freq × English legality	2.59846	0.56704	4.583	***

Table A.6.: Summary of the model featuring L1 legality instead of L1 frequency
Significance codes: *** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1
Formula: `error~logFreqDE*legalityEN + son.vio + ons.intensity + accNF + NADdiff + (logFreqDE*legalityEN + son.vio + accNF + NADdiff|subjID) + (1|onset.targ/stimulus)`

B. Production experiment

B.1. Stimulus material

Table B.1.: Stimulus items in the production experiment; order of syllables within a stimulus corresponds to List A; syllable 1 and syllable 2 are reversed for all test stimuli in List B

CCs	stimulus	status	rd[a]
ʃt–tʃ	ʃtɔm tʃɔm	test	2
	ʃtæː tʃaː	test	1
	ʃtɛn tʃɛr	test	1
	ʃtœf tʃaf	test	2
	tʃʊx ʃtɔr	test	1
	tʃap ʃtɪn	test	3
	tʃʊn ʃtʊl	test	2
	tʃʊl ʃtʏl	test	2
ʃt–ʃn	ʃtaːn ʃnaːt	test	2
	ʃtɪŋ ʃnʊk	test	1
	ʃtuː ʃnøː	test	1
	ʃtʊx ʃnʊŋ	test	2
	ʃnɔɪ̯p ʃtɔɪ̯m	test	2
	ʃnyː ʃtoː	test	1
	ʃnɪr ʃtɪr	test	1
	ʃnɪf ʃtɛf	test	2
ʃp–ʃm	ʃpoː ʃmaʊ̯	test	1
	ʃpɪŋ ʃmɪr	test	1
	ʃpʏç ʃmʊx	test	2
	ʃpaf ʃmaf	test	2
	ʃmɛl ʃpɛl	test	2

continued on next column

[a] rated difficulty; only used for pseudo-randomisation of stimulus order

Table B.1 *(continued)*

CCs	stimulus	status	rd
	ʃmoːt ʃpaːk	test	2
	ʃmɔx ʃpɔf	test	2
	ʃmaŋ ʃpʊŋ	test	2
tr–kr	traː kraː	test	1
	trɛl krɛŋ	test	2
	trʊs krʏs	test	2
	tral kreːf	test	1
	kruː troː	test	1
	krɔf trɔʃ	test	2
	kroːl tryːm	test	2
	kriːs triːs	test	1
ʃl–fl	ʃluː fluː	test	1
	ʃloːm flaʊ̯n	test	2
	ʃlɛr flɛm	test	1
	ʃleːç fløːç	test	1
	flɛm ʃlɔɪ̯n	test	2
	flaːk ʃlaːx	test	2
	flœp ʃlʏt	test	2
	fliːp ʃliːp	test	1
fl–sl	flɔɪ̯ slɔɪ̯	test	1
	flʊp slʊʃ	test	1
	fliːm sloːn	test	1
	flaŋ slʊŋ	test	1
	sluː fløː	test	1
	slɪt flɛp	test	2
	slɔk flɔk	test	1
	slɛt flɛç	test	1
pl–ps	plɪf psɪç	test	2
	plɔɪ̯ç psɔɪ̯n	test	2
	pluːx psoːx	test	1

continued on next column

Table B.1 *(continued)*

CCs	stimulus	status	rd
	plaʃ psʊʃ	test	2
	psoː ploː	test	1
	psɪn plœn	test	2
	psaːn plɪm	test	1
	psøːf pløːs	test	2
ts–ks	tsɔɪ̯ ksaɪ̯	test	2
	tsyːl ksyːl	test	2
	tsɛm ksaɪ̯n	test	2
	tsʊf ksʊm	test	2
	ksaː tsaː	test	2
	ksʏç tsɪf	test	2
	ksɛl tsɪr	test	2
	kseːl tsoːr	test	2
ks–sk	ksuː skuː	test	2
	ksyːt skɔr	test	1
	ksan skʊm	test	3
	ksɛl skɛf	test	3
	skaʊ̯ ksɔɪ̯	test	2
	skɪr ksɛp	test	2
	skoːn ksoːf	test	3
	skeːl kseːl	test	2
ps–sp	psaɪ̯ spɔɪ̯	test	2
	psøːt spøːl	test	3
	psaf spɔf	test	3
	psɛl spɛr	test	2
	spaʊ̯ psaʊ̯	test	3
	spɛf psɪç	test	2
	spuːx psuːk	test	3
	spʊm psɔm	test	3
—	ʃoː riːl	filler	1

continued on next column

Table B.1 *(continued)*

CCs	stimulus	status	rd
	gøː zuː	filler	1
	hiːt beː	filler	1
	kaɪ̯ taː	filler	1
	biː loː	filler	1
	koː rɪç	filler	1
	raɪ̯ goː	filler	1
	køː riː	filler	1
	buː reːl	filler	1
	tɛl vuː	filler	1
	doː tiː	filler	1
	faɪ̯ dak	filler	1
	kiː faɪ̯	filler	1
	guː vaː	filler	1
	ka fɛŋ	filler	1
	jɔs paː	filler	1
	pax raː	filler	1
	fuː paːk	filler	1
	maʃ guː	filler	1
	liːk paː	filler	1
	gʊp giː	filler	1
	fɛm raː	filler	1
	luː baː	filler	1
	vɔp deː	filler	1
	gɔl riː	filler	1
	bɛŋ ʃoː	filler	1
	koː vyː	filler	1
	ʃʏl doː	filler	1
	hɛn faː	filler	1
	feː daŋ	filler	1
	fɛs kiː	filler	1

continued on next column

Table B.1 *(continued)*

CCs	stimulus	status	rd
	kʊŋ ʃeː	filler	1
	pøː tam	filler	1
	hiː miː	filler	1
	tɛr ʃaː	filler	1
	hoː faɪ̯	filler	1
	gɔl ʃoː	filler	1
	dʊf røː	filler	1
	nuːp foː	filler	1
	gɔɪ̯ fɛr	filler	1
	lɔɪ̯ han	filler	1
	ɔʃ biː	filler	1
	hoː laːp	filler	1
	lɛç tam	filler	1
	tɪn pal	filler	1
	voːt lyː	filler	1
	giː pɔŋ	filler	1
	raːl deː	filler	1
	kɔr ʃøːp	filler	1
	tʊn fɛ3k	filler	1

B.2. Questionnaire

1. Alter:
2. Geschlecht: □ weiblich □ männlich
3. Händigkeit: □ RechtshänderIn □ LinkshänderIn
4. Studienfächer (auch frühere):
5. Geburtsort:
6. weitere Wohnorte (Geben Sie bitte auch an, wie lange Sie in etwa an dem jeweiligen Ort gelebt haben):
7. Kommen oder kamen Sie regelmäßig mit einem oder mehreren deutschen Dialekten in Berührung (z.B. durch Eltern, FreundInnen, PartnerIn)? □ nein □ ja
   Falls ja, mit welchem/n und wie oft?
8. Sprechen Sie weitere Sprachen? □ nein □ ja
   Falls ja, welche und wie gut (siehe Skala)?
   5 = nahezu muttersprachlich
   4 = sehr gut
   3 = gut
   2 = mittelmäßig
   1 = nicht besonders gut
9. Spielen Sie ein Instrument oder sind Sie anderweitig musikalisch aktiv?
   □ nein □ ja (welches/wie oft/seit wann)
10. Haben Sie Hörprobleme? □ nein □ ja und zwar
11. Waren Sie mal in logopädischer Behandlung oder wurden bei Ihnen Sprachentwicklungsstörungen diagnostiziert?
    □ nein □ ja und zwar

Vielen Dank!

B.3. Statistical models and additional effect plots

predictor	estimate	SE	z value	
(Intercept)	-2.57340	0.22735	-11.319	***
log cluster frequency (type)	-0.38849	0.16084	-2.415	*
cons. metathesis	-0.98175	0.09553	-10.277	***
SSP violation	0.17967	0.19148	0.938	
coda in prev.	-1.27808	0.07364	-17.356	***
coda identical no	-0.56560	0.07459	-7.583	***
coda identical yes	-0.61659	0.08247	-7.476	***
log cluster freq × metathesis	-0.21053	0.05708	-3.688	**

Table B.2.: Model output of the best-fitting model (data subset excluding repetition errors)
Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$
Formula: `error ~logFreq*metathesis + son.vio + complex + coda.id + (logFreq*metathesis + son.vio|subjID) + (1|onset.targ/stimulus)`

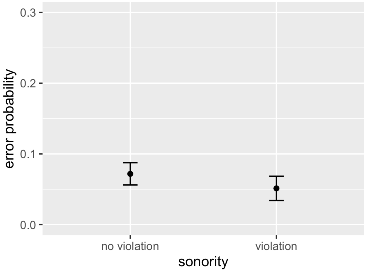


Figure B.1.: Effect of SSP violation

predictor	estimate	SE	z value	
(Intercept)	-3.49814	0.20967	-16.684	***
log cluster frequency (type)	-0.23270	0.07722	-3.014	**
consonant metathesis	-0.92693	0.09308	-9.958	***
sonority distance −1	0.76413	0.20873	3.661	***
sonority distance 1	3.43602	0.47319	7.261	***
sonority distance 2	2.43305	0.52566	4.629	***
generalised frequency	0.49019	0.08364	5.861	***
coda in prev.	-1.28497	0.07367	-17.442	***
coda identical no	-0.56501	0.07545	-7.489	***
coda identical yes	-0.62996	0.08314	-7.577	***
log cluster freq × metathesis	-0.17535	0.05307	-3.304	***

Table B.3.: Output of the model including generalised frequency
Significance codes: *** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1
Formula: error ~logFreq*metathesis + genFreq + son.dist + complex + coda.id + (logFreq*metathesis + son.dist|subjID) + (1|onset.targ/stimulus)

predictor	estimate	SE	z value	
(Intercept)	-4.14757	0.29791	-13.922	***
summed phoneme frequency (type)	-0.67909	0.14086	-4.821	***
cons. metathesis	2.25378	0.19199	11.739	***
sonority distance 1 (ref. level: -1)	1.07996	0.27949	3.864	***
sonority distance 2	0.21437	0.36090	0.594	
sonority distance 3	0.46361	0.38584	1.202	
coda in prev.	-1.28034	0.07347	-17.426	***
coda identical yes (ref. level: no)	-0.55614	0.07450	-7.465	***
coda identical no coda	0.64717	0.08246	-7.848	***
summed phon. freq*metathesis	0.62867	0.13457	4.672	***

Table B.4.: Output of the model with summed phoneme frequency
Significance codes: *** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1
Formula: error ~phonFreq*metathesis + son.dist + complex + coda.id + (phonFreq*metathesis + son.dist|subjID) + (1|onset.targ/stimulus)

predictor	estimate	SE	z value	
(Intercept)	-5.0092	0.3485	-14.371	***
log frequency difference	0.2255	0.2472	0.912	
sonority improvement	0.1196	0.2137	0.560	

Table B.5.: Summary of the model predicting internal substitution rate (dataset excluding metathesis pairs)
Significance codes: *** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1
Formula: `part.prod ~logFreqDiff + son.improve + (logFreqDiff + son.improve|subj.no) + (1|ons.targ/syllable)`

consonant cluster	non-metathesis	metathesis	t statistic
ʃt	M = 0.97, SD = 1.8	M = 5.97, SD = 3.6	*t*(57) = -7.8, *p* < .001
ks	M = 4.76, SD = 3.7	M = 12.77, SD = 5.4	*t*(68) = -7.6, *p* < .001
ps	M = 5.77, SD = 4.2	M = 9.44, SD = 5.1	*t*(73) = -3.5, *p* < .001

Table B.6.: Results of Welch's two-sample t-tests (one-tailed) for error rates of the same clusters in metathesis pairs vs. non-metathesis pairs; hypothesis: error rate in non-metathesis pairs is smaller

B.4. Confusion matrix

Table B.7.: Confusion Matrix
Rows show target clusters, columns productions
(Since a skewed distribution of NAs over clusters made interpretation of percentages difficult, it was decided to present raw numbers instead.)

	ts	ʃt	ʃp	tr	kr	ʃl	fl	ʃm	pl	ʃn	sk	ps	sl	tʃ	ks	sp	add.	C1 del.	C2 del.	other
ts	793	0	0	0	0	0	0	0	0	0	3	2	0	0	51	0	24	29	2	6
ʃt	0	2000	1	0	0	1	0	0	0	28	0	0	0	73	0	0	101	2	32	33
ʃp	0	0	1148	0	0	0	0	24	0	1	0	0	0	0	0	1	8	3	0	15
tr	2	1	0	1181	6	0	0	0	0	0	0	0	0	0	0	0	7	0	14	3
kr	0	0	0	10	1157	0	0	0	0	0	0	0	0	0	0	0	21	0	18	5
ʃl	0	1	0	0	0	1132	13	1	0	2	0	0	0	0	0	0	30	1	12	4
fl	0	0	0	0	0	15	2240	0	0	0	0	0	3	0	0	0	27	0	22	8
ʃm	0	0	41	0	0	0	0	1115	0	0	0	0	0	0	0	0	8	22	0	10
pl	1	0	0	0	0	1	1	0	1029	0	0	4	0	0	0	0	22	0	3	19
ʃn	0	48	0	0	0	2	0	0	0	1143	0	0	0	0	0	0	11	1	2	3
sk	5	0	0	0	0	0	0	0	0	0	826	0	0	0	49	1	75	1	32	5
ps	60	0	2	0	0	0	0	0	46	0	1	1367	2	0	1	172	222	41	6	40
sl	2	0	0	0	0	9	5	0	0	0	0	0	1044	0	0	0	32	1	14	9
tʃ	1	193	1	0	0	0	0	0	0	0	0	0	0	668	0	0	98	61	2	13
ks	55	0	0	0	0	0	0	0	0	0	365	2	0	0	1231	1	207	26	7	16
sp	5	0	4	0	0	0	0	0	0	0	2	88	0	0	0	676	109	5	18	31

Bibliography

Adriaans, F., & Kager, R. W. J. (2010). Adding generalization to statistical learning : The induction of phonotactics from continuous speech. *Journal of Memory and Language*, *62*(3), 311–331. https://doi.org/10.1016/j.jml.2009.11.007

Aichert, I., & Ziegler, W. (2004). Syllable frequency and syllable structure in apraxia of speech. *Brain and Language*, *88*(1), 148–159. https://doi.org/10.1016/S0093-934X(03)00296-7

Albright, A. (2007a). *Gradient phonological acceptability as a grammatical effect*, Dept. of Lingustics, MIT. https://doi.org/https://doi.org/10.1017S

Albright, A. (2007b). Natural classes are not enough: Biased generalization in novel onset clusters. *15th Manchester Phonology Meeting*, 1–30. http://web.mit.edu/albright/www/papers/Albright-BiasedGeneralization.pdf

Albright, A. (2009). Feature-based generalisation as a source of gradient acceptability. *Phonology*, *26*(1), 9–41. https://doi.org/10.1017/S0952675709001705

Albright, A. (2012). Modeling morphological productivity with the Minimal Generalization Learner. *Data-Rich Approaches to English Morphology: From Corpora and Experiments to Theory and Back (DRAEM).*

Alderete, J., & Tupper, P. (2018). Phonological regularity, perceptual biases, and the role of phonotactics in speech error analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, *9*(5), e1466. https://doi.org/10.1002/wcs.1466

Ali, E. M. T., & Van Heuven, V. J. (2009). Segmental analysis of speech intelligibility problems among Sudanese listeners of English. *English as International Language Journal*, *4*(August), 129–165.

Altenberg, E. P. (2005). The judgment, perception, and production of consonant clusters in a second language. *IRAL - International Review of Applied Linguistics in Language Teachings*, *43*(1), 53–80. https://doi.org/10.1515/iral.2005.43.1.53

Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*, 239–273.

Andrews, S. (1992). Frequency and Neighborhood Effects on Lexical Access: Lexical Similarity or Orthographic Redundancy? *Journal of Experimental Psychology*, *18*(2), 234–254.

Archer, S. L., & Curtin, S. (2011). Perceiving onset clusters in infancy. *Infant Behavior and Development*, *34*(4), 534–540. https://doi.org/10.1016/j.infbeh.2011.07.001

Arnon, I. (2015). What can frequency effects tell us about the building blocks and mechanisms of language learning? *Journal of Child Language*, *42*(2), 274–277. https://doi.org/10.1017/S0305000914000610

Arons, B. (1992). A Review of The Cocktail Party Effect. *Journal of the American Voice I/O Society*, *12*, 35–50.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324.

Auer Jr., E. T., & Luce, P. A. (2005). Probabilistic Phonotactics in Spoken Word Recognition. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of speech perception* (pp. 610–630). Blackwell.

Baars, B. J. (1980). The competing plans hypothesis: An heuristic viewpoint on the causes of errors in speech. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of frieda goldman-eisler* (pp. 39–49). De Gruyter Mouton. https://doi.org/10.1515/9783110816570.39

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). CELEX2 LDC96L14.

Barlow, J. (2005). Sonority effects in the production of consonant clusters by Spanishspeaking children. *Selected proceedings from the 6th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*, 1–14. http://www.lingref.com/cpp/casp/6/paper1121.pdf

Baroni, A. (2013). Element Theory and the Magic of /s/.

Baroni, A. (2014). On the importance of being noticed: the role of acoustic salience in phonotactics (and casual speech). *Language Sciences*, *46*, 18–36. https://doi.org/10.1016/j.langsci.2014.06.004

Bastiaanse, R., Gilbers, D., & van der Linde, K. (1994). Sonority substitutions in Broca's and Conduction Aphasia. *Journal of Neurolinguistics*, *8*(4), 247–255.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious Mixed Models. *arXiv preprint arXiv:1506.04967*, 1–27. https://doi.org/arXiv:1506.04967

Baumann, A., & Ritt, N. (2018). The basic reproductive ratio as a link between acquisition and change in phonotactics. *Cognition*, *176*(March 2017), 174–183. https://doi.org/10.1016/j.cognition.2018.03.005

Behrens, H., & Pfänder, S. (Eds.). (2016). *Experience counts: Frequency effects in language (Vol. 54)*. Walter de Gruyter GmbH & Co KG.

Bell, A., Brenier, J. M., Gregory, M. L., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–111. https://doi.org/10.1016/j.jml.2008.06.003

Bellanova, M. (2016). *Development of a Logatome Test for the Evaluation of Signal Processing Algorithms in Hearing Aids on a Microscopic Level* (Doctoral dissertation). Friedrich-Alexander-Universität Erlangen-Nürnberg.

Benkí, J. R. (2002). Effects of Signal-Independent Factors in Speech Perception. *Annual Meeting of the Berkeley Linguistics Society*, *28*(1), 63. https://doi.org/10.3765/bls.v28i1.3823

Benkí, J. R. (2003). Analysis of english nonsense syllable recognition in noise. *Phonetica*, *60*(2), 129–157. https://doi.org/10.1159/000071450

Berent, I. (2016). Commentary: "an evaluation of universal grammar and the phonological mind"-UG is still a viable hypothesis. *Frontiers in Psychology*, *7*(JUL), 1–13. https://doi.org/10.3389/fpsyg.2016.01029

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, *104*, 591–630. https://doi.org/10.1016/j.cognition.2006.05.015

Berg, T. (1989). Intersegmental cohesiveness*. *Folia Linguistica*, *23*(3-4), 245–280. https://doi.org/10.1515/flin.1989.23.3-4.245

Berg, T. (2014). On the Relationship between Type and Token Frequency. *Journal of Quantitative Linguistics*, *21*(3), 199–222. https://doi.org/10.1080/09296174.2014.911505

Best, C. T. (1994). The Emergence of Native-Language Phonological Influences in Infants: A Perceptual Assimiliation Model. *The Development of Speech Perception*, *167*(224), 233–277. https://doi.org/10.7551/mitpress/2387.003.0011

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. In O.-S. Bohn & Murray J. Munro (Eds.), *Language experience in second language speech learning: In honor of james emil flege* (pp. 13–34). John Benjamins Publishing. https://doi.org/10.1075/lllt.17.07bes

Bislick, L., & Hula, W. D. (2019). Perceptual characteristics of consonant production in apraxia of speech and aphasia. *American Journal of Speech-Language Pathology*, *28*(4), 1411–1431. https://doi.org/10.1044/2019_AJSLP-18-0169

Bloch, D. (2007). *On Memory and Recollection. Text, Translation, Interpretation, and Reception in Western Scholasticism* (D. Bloch, Ed.). Brill.

Blumenfeld, H. K., & Marian, V. (2013). Parallel language activation and cognitive control during spoken word recognition in bilinguals. *Journal of Cognitive Psychology*, *25*(5), 547–567. https://doi.org/10.1080/20445911.2013.812093

Blumenthal-Dramé, A. (2012). *Entrenchment in usage-based theories: What corpus data do and do not reveal about the mind* (Vol. 83). Walter de Gruyter.

Blumenthal-Dramé, A. (2017). Entrenchment from a psycholinguistic and neurolinguistic perspective. In H.-J. Schmid (Ed.), *Language and the human lifespan series. entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge* (pp. 129–152). De Gruyter Mouton.

Boersma, P. (1998). *Functional Phonology* (Doctoral dissertation). Universiteit van Amsterdam.

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer. http://www.praat.org

Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *NeuroImage*, *32*(2), 821–841. https://doi.org/10.1016/j.neuroimage.2006.04.173

Boll-Avetisyan, N. (2011). Second Language Probabilistic Phonotactics and Structural Knowledge in Short-Term Memory Recognition. *BUCLD 35 Proceedings*, *35*(April), 60–72.

Boll-Avetisyan, N. (2012). *Phonotactics and Its Acquisition, Representation, and Use - An Experimental-Phonological Study* (Doctoral dissertation). Universiteit Utrecht.

Bombien, L., Mooshammer, C., Hoole, P., & Kühnert, B. (2010). Prosodic and segmental effects on EPG contact patterns of word-initial German clusters. *Journal of Phonetics*, *38*(3), 388–403. https://doi.org/10.1016/j.wocn.2010.03.003

Bond, Z. S. (1971). *Units in Speech Perception* (Doctoral dissertation). Ohio State University.

Boomer, D. S., & Laver, J. D. M. (1989). Slips of the Tongue. *Philosophical Psychology*, *2*(2), 203–222. https://doi.org/10.1080/09515088908572972

Borsky, S., Tuller, B., & Shapiro, L. P. (1998). "How to milk a coat:" The effects of semantic and acoustic information on phoneme categorization. *The Journal of the Acoustical Society of America*, *103*(5), 2670–2676. https://doi.org/10.1121/1.422787

Bose, A., van Lieshout, P., & Square, P. a. (2007). Word frequency and bigram frequency effects on linguistic processing and speech motor performance in individuals with aphasia and normal speakers. *Journal of Neurolinguistics*, *20*(1), 65–88. https://doi.org/10.1016/j.jneuroling.2006.05.001

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners. *The Journal of the Acoustical Society of America*, *29*(10), 1883–1889.

Bregman, A. S., & McAdams, S. (1994). Auditory Scene Analysis: The Perceptual Organization of Sound. *The Journal of the Acoustical Society of America*, *95*(2), 1177–1178. https://doi.org/10.1121/1.408434

Brendel, B., Ziegler, W., Erb, M., Riecker, A., & Ackermann, H. (2008). Does our Brain House a "Mental Syllabary"? An fMRI Study. *Proceedings the 8th International Seminar on Speech Production (ISSP)*, 73–76.

Broselow, E., & Finer, D. (1991). Parameter setting in second language phonology and syntax. *Second Language Research*, *7*(1), 35–59. https://doi.org/10.1177/026765839100-700102

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, *3*(1986), 219–252. https://doi.org/10.1017/s0952675700000658

Brown, R. W., & Hildum, D. C. (1956). Expectancy and the perception of syllables. *Language*, *32*(3), 411–419.

Brunner, J., Geng, C., Sotiropoulou, S., & Gafos, A. I. (2014). Timing of German onset and word boundary clusters. *Laboratory Phonology*, *5*(4), 403–454. https://doi.org/10.1515/lp-2014-0014

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect. *Experimental psychology*, *58*(5), 412–24. https://doi.org/10.1027/1618-3169/a000123

Buchwald, A. (2009). Minimizing and optimizing structure in phonology: Evidence from aphasia. *Lingua*, *119*(10), 1380–1395. https://doi.org/10.1016/j.lingua.2007.11.015

Buchwald, A. (2014). Phonetic Processing. In V. S. Ferreira, M. Goldrick, & M. Miozzo (Eds.), *Oxford handbook of language production* (pp. 245–258). Oxford University Press.

Bürki, A., Cheneval, P. P., & Laganaro, M. (2015). Do speakers have access to a mental syllabary? ERP comparison of high frequency and novel syllable production. *Brain and Language*, *150*, 90–102. https://doi.org/10.1016/j.bandl.2015.08.006

Bußmann, H. (1983). *Lexikon der Sprachwissenschaft*. Alfred Kröner.

Buz, E., & Jaeger, T. F. (2016). The (in)dependence of articulation and lexical planning during isolated word production. *Language, Cognition and Neuroscience*, *31*(3), 404–424. https://doi.org/10.1080/23273798.2015.1105984

Bybee, J. L. (1999). Usage-based phonology. In M. Darnell, E. A. Moravcsik, M. Noonan, F. J. Newmeyer, & K. Wheatley (Eds.), *Functionalism and formalism in linguistics* (pp. 211–242). John Benjamins. https://www.unm.edu/%7B~%7Djbybee/downloads/Bybee1999UsageBasedPhonology.pdf

Bybee, J. L. (2002). Phonological Evidence for Exemplar Storage of Multiword Sequences. *Studies in Second Language Acquisition*, *24*(2), 215–221. https://doi.org/10.1017/s0272263102002061

Bybee, J. L. (2010). *Language, usage and cognition.* https://doi.org/10.1353/lan.2011.0082

Bybee, J. L., & McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *Linguistic Review*, *22*(2-4), 381–410. https://doi.org/10.1515/tlir.2005.22.2-4.381

Byrd, D., & Choi, S. (2010). At the juncture of prosody, phonology, and phonetics - The interaction of phrasal and syllable structure in shaping the timing of consonant gestures. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory phonology* (pp. 31–60). De Gruyter Mouton.

Carlson, M. T. (2018). Now you hear it, now you don't: Malleable illusory vowel effects in Spanish-English bilinguals. *Bilingualism: Language and Cognition*, *22*(5), 1101–1122. https://doi.org/10.1017/S136672891800086X

Carlson, M. T., Goldrick, M., Blasingame, M., & Fink, A. (2016). Navigating conflicting phonotactic constraints in bilingual speech perception. *Bilingualism: Language and Cognition*, *19*(5), 1–16. https://doi.org/10.1017/S1366728915000334

Carré, R., Bourdeau, M., & Tubach, J.-P. (1995). Vowel-vowel production: The distinctive region model (DRM) and vowel harmony. *Phonetica*, *52*(3), 205–214.

Carreiras, M., & Perea, M. (2004). Naming pseudowords in Spanish: Effects of syllable frequency. *Brain and Language*, *90*(1-3), 393–400. https://doi.org/10.1016/j.bandl.2003.12.003

Cason, N., Astésano, C., & Schön, D. (2015). Bridging music and speech rhythm: Rhythmic priming and audio-motor training affect speech perception. *Acta Psychologica*, *155*, 43–50. https://doi.org/10.1016/j.actpsy.2014.12.002

Chang, S. S., Plauché, M. C., & Ohala, J. J. (2001). Markedness and consonant confusion asymmetries. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 79–101). Academic Press.

Cheng, X., Schafer, G., & Riddell, P. M. (2014). Immediate Auditory Repetition of Words and Nonwords : An ERP Study of Lexical and Sublexical Processing. *PLoS One*, *9*(3). https://doi.org/10.1371/journal.pone.0091988

Chikamatsu, N. (2006). Developmental word recognition: A study of L1 english readers of L2 Japanese. *Modern Language Journal*, *90*(1), 67–85. https://doi.org/10.1111/j.1540-4781.2006.00385.x

Cholin, J., Dell, G. S., & Levelt, W. J. M. (2011). Planning and articulation in incremental word production: syllable-frequency effects in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 109–122. https://doi.org/10.1037/a0021322

Cholin, J., & Levelt, W. J. M. (2009). Effects of syllable preparation and syllable frequency in speech production: Further evidence for syllabic units at a post-lexical level. *Language and Cognitive Processes*, *24*(5), 662–684. https://doi.org/10.1080/01690960802348852

Cholin, J., Levelt, W. J. M., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition*, *99*(2), 205–235. https://doi.org/10.1016/j.cognition.2005.01.009

Christman, S. S. (1992). Uncovering phonological regularity in neologisms: Contributions of sonority theory. *Clinical Linguistics and Phonetics*, *6*(3), 219–247. https://doi.org/10.3109/02699209208985532

Christman, S. S. (1994). Target-Related Neologism Formation in Jargonaphasia. *Brain and Language*, *46*, 109–128.

Claxton, G. L. (1974). Initial consonant groups function as units in word production. *Language and Speech*, *17*(3), 271–277. https://doi.org/10.1177/002383097401700305

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology* (pp. 283–333). Cambridge University Press.

Clements, G. N. (2009). Does sonority have a phonetic basis? Comments on the chapter by Vaux. *Contemporary Views on Architec-*

ture and Representations in Phonological Theory, 165–175. https://halshs.archives-ouvertes.fr/halshs-00182675

Cluff, M., & Luce, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 551–563.

Code, C., & Ball, M. J. (1994). Syllabification in aphasic recurring utterances: contributions of sonority theory. *Journal of Neurolinguistics*, *8*(4), 257–265. https://doi.org/10.1016/0911-6044(94)90012-4

Cohen, S. P., Tucker, G. R., & Lambert, W. E. (1967). The comparative skills of monolinguals and bilinguals in perceiving phoneme sequences. *Language and Speech*, *10*(3), 159–168.

Cole, J. S., & Iskarous, K. (2001). Effects of Vowel Context On Consonant Place Identification : Implications for a Theory of Phonologization. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 103–122). Academic Press San Diego.

Coleman, J., & Pierrehumbert, J. B. (1997). Stochastic phonological grammars and acceptability. *Computational phonology Third meeting of the ACL special interest group in computational phonology*, 8. https://doi.org/10.3109/13682820109177934

Connell, L., & Lynott, D. (2012). Strength of Perceptual Experience Predicts Word Processing Performance Better than Concreteness or Imageability. *UC Merced Proceedings of the Annual Meeting of the Cognitive Science Society*, (34), 34. https://cloudfront.escholarship.org/dist/prd/content/qt01b1f668/qt01b1f668.pdf

Connine, C. M., Blasko, D., & Titone, D. (1993). Do the Beginnings of Spoken Words Have a Special Status in Auditory Word Recognition? *Journal of Memory & Language*, *32*, 193–210.

Conrad, M., Carreiras, M., & Jacobs, A. M. (2008). Contrasting effects of token and type syllable frequency in lexical decision. *Language and Cognitive Processes*, *23*(2), 296–326. https://doi.org/10.1080/01690960701571570

Cortese, M. J., & Schock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *Quarterly Journal of Experimental Psychology*, *66*(5), 946–972. https://doi.org/10.1080/17470218.2012.722660

Crompton, A. (1981). Syllables and segments in speech production. *Linguistics*, *19*(7-8), 663–716. https://doi.org/10.1515/9783110828306.109

Croot, K., & Rastle, K. (2004). Is there a syllabary containing stored articulatory plans for speech production in English. *Proceedings of the 10th Australian International Conference on Speech Science and Technology*, (December), 376–381. http://www.psych.usyd.edu.au/staff/karenc/SST04Croot%7B%5C%%7D26Rastle4.pdf

Cutler, A. (1981). The reliability of speech error data. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 561–582). de Gruyter.

Cutler, A. (2012). *Native Listening. Language Experience and the Recognition of Spoken Words*. MIT Press.

Cutler, A., Butterfield, S., & Williams, J. N. (1987). The Perceptual Integrity of Syllabic Onsets. *Journal of Memory and Language*, *26*, 406–418.

Cutler, A., & Clifton, C. (2000). Comprehending spoken language: a blueprint of the listener. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (Paperback, pp. 123–165).

Cutler, A., Garcia Lecumberri, M. L., & Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: effects of local context. *The Journal of the Acoustical Society of America*, *124*(2), 1264–1268. https://doi.org/10.1121/1.2946707

Dąbrowska, E. (2012). Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism*, *2*(3), 219–253.

Daland, R., Hayes, B., White, J., Garellek, M., Davis, A., & Norrmann, I. (2011). *Explaining sonority projection effects* (Doctoral dissertation No. *2*). https://doi.org/10.1017/S0952675711000145

Davidson, L. (2011). Phonetic, Phonemic, and Phonological Factors in Cross-Language Discrimination of Phonotactic Contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 270–282. https://doi.org/10.1037/a0020988

Davidson, L., & Shaw, J. A. (2012). Sources of illusion in consonant cluster perception. *Journal of Phonetics*, *40*(2), 234–248. https://doi.org/10.1016/j.wocn.2011.11.005

De Smet, H. (2016). How gradual change progresses: The interaction between convention and innovation. *Language Variation and Change*, *28*(1), 83–102. https://doi.org/10.1017/S0954394515000186

Decoene, S. (1993). Testing the speech unit hypothesis with the primed matching task: Phoneme categories are perceptually basic. *Perception & Psychophysics*, *53*(6), 601–616. https://doi.org/10.3758/BF03211737

Del Prado Martín, F. M., Ernestus, M., & Baayen, R. H. (2004). Do type and token effects reflect different mechanisms? Connectionist modeling of Dutch past-tense formation and final devoicing. *Brain and Language*, *90*(1-3), 287–298. https://doi.org/10.1016/j.bandl.2003.12.002

Dell, G. S. (1986). A Spreading-Activation Theory of Retrieval in Sentence Production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, *27*(2), 124–142. https://doi.org/10.1016/0749-596X(88)90070-8

Dell, G. S. (2000). Commentary: Counting, connectionism, and lexical representation. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in* *laboratory phonology v: Acquisition and the lexicon* (pp. 335–348). Cambridge University Press.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, *104*(1), 123–147. https://doi.org/10.1037/0033-295X.104.1.123

Dell, G. S., Juliano, C., & Govindjee, A. (1993a). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, *17*(2), 149–195. https://doi.org/10.1016/0364-0213(93)90010-6

Dell, G. S., Juliano, C., & Govindjee, A. (1993b). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, *17*(2), 149–195. https://doi.org/10.1016/0364-0213(93)90010-6

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech Errors, Phonotactic Constraints, and Implicit Learning: A Study of the Role of Experience in Language Production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1355–1367. https://doi.org/10.1037/0278-7393.26.6.1355

Deschamps, I., Baum, S. R., & Gracco, V. L. (2015). Phonological processing in speech perception: What do sonority differences tell us? *Brain and Language*, *149*, 77–83. https://doi.org/10.1016/j.bandl.2015.06.008

Diessel, H. (2003). Review: Frequency and the Emergence of Linguistic Structure. *Linguistics*, *39*(1), 167–172.

Diessel, H. (2007). Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology*, *25*(2), 108–127. https://doi.org/10.1016/j.newideapsych.2007.02.002

Diessel, H. (2017). Usage-Based Linguistics. In M. Aronoff (Ed.), *Oxford research encyclopedia of linguistics* (pp. 1–26). Oxford University Press. https://doi.org/10.1093/obo/9780199772810-0068

Dunn, A. L., & Fox Tree, J. E. (2014). More on language mode. *International Journal of Bilingualism*, *18*(6), 605–613. https://doi.org/10.1177/1367006912454509

Dupoux, E., Hirose, Y., Kakehi, K., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1568–1578. https://doi.org/10.1037/0096-1523.25.6.1568

Dupoux, E., Pallier, C., Kakehi, K., & Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes*, *16*, 491–505. https://doi.org/10.1080/01690960143000191

Dziubalska-Kołaczyk, K. (2001). *Beats-and-Binding Phonology*. Peter Lang.

Dziubalska-Kołaczyk, K. (2007). Natural Phonology: Universal principles for the study of language. In J. Trouvain & W. Barry (Eds.), *Proceedings of the 16th international congress of phonetic sciences* (pp. 71–75).

Dziubalska-Kołaczyk, K. (2009). NP Extension: B&B Phonotactics. *Poznań Studies in Contemporary Linguistics*, *45*(1). https://doi.org/10.2478/v10010-009-0011-9

Dziubalska-Kołaczyk, K. (2014). Explaining phonotactics using NAD. *Language Sciences*, *46*, 6–17. https://doi.org/10.1016/j.langsci.2014.06.003

Dziubalska-Kołaczyk, K. (2015). Are frequent, early and easy clusters also unmarked? *Rivista di Linguistica*, *27*(1), 29–43.

Dziubalska-Kołaczyk, K. (2019). On the structure, survival and change of consonant clusters. *Folia Linguistica*, *40*(1), 107–127. https://doi.org/10.1515/flih-2019-0006

Dziubalska-Kołaczyk, K., Pietrala, D., & Aperlinski, G. (n.d.). The NAD Phonotactic Calculator – an online tool to calculate cluster preference in English, Polish and other languages. Retrieved August 3, 2018, from http://wa.amu.edu.pl/nadcalc/

Edwards, J., Beckman, M. E., & Munson, B. (2004). Vocabulary size and phonotactic production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, *47*(2), 421–436. https://doi.org/10.1044/1092-4388(2004/034)

Eimas, P. D., & Corbit, J. D. (1973). Selective Adaptation of Linguistic Feature Detectors. *Cognitive Psychology*, *4*, 99–109.

Elert, C.-C. (1970). *Ljud och ord i svenskan*. Almqvist & Wiksell.

Elexiko. (2003). http://www.owid.de/wb/elexiko/start.html

Ellis, N. C. (2002). Frequency Effects in Language Processing. *Studies in Second Language Acquisition*, *24*(2), 143–188. https://doi.org/10.1093/nq/s6-VII.162.89-a

Ellis, N. C., & Collins, L. (2009). Input and second language acquisition: The roles of frequency, form, and function introduction to the special issue. *Modern Language Journal*, *93*(3), 329–335. https://doi.org/10.1111/j.1540-4781.2009.00893.x

Elman, J. L., Diehl, R. L., & Buchwald, S. E. (1977). Perceptual switching in bilinguals. *The Journal of the Acoustical Society of America*, *62*(4), 971–974. https://doi.org/10.1121/1.381591

Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, *142*, 27–41. https://doi.org/10.1016/j.lingua.2012.12.006

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*(4), 551–585.

Everett, D. L. (2016). An Evaluation of Universal Grammar and the Phonological Mind. *Frontiers in Psychology*, *7*(February), 1–10. https://doi.org/10.3389/fpsyg.2016.00015

Felty, R. A., Buchwald, A., Gruenenfelder, T. M., & Pisoni, D. B. (2013). Misperceptions of spoken words: Data from a random sample of American English words. *The Journal of the Acoustical Society of America*, *134*(1), 572–85. https://doi.org/10.1121/1.4809540

Foss, D. J., & Blank, M. A. (1980). Identifying the Speech Codes. *Cognitive Psychology*, *12*, 1–31.

Foss, D. J., & Swinney, D. A. (1973). On the psychological reality of the phoneme: Perception, identification, and consciousness. *Journal of Verbal Learning and Verbal Behavior*, *12*(3), 246–257. https://doi.org/10.1016/S0022-5371(73)80069-6

Frauenfelder, U. H., Baayen, R. H., & Hellwig, F. (1993). Neighborhood Density and Frequency Across Languages and Modalities. *Journal of Memory and Language*, *32*(6), 781–804.

Frauenfelder, U. H., & Segui, J. (1989). Phoneme monitoring and lexical processing: evidence for associative context effects. *Memory & Cognition*, *17*(2), 134–140.

Freeman, M. R., Blumenfeld, H. K., & Marian, V. (2016). Phonotactic constraints are activated across languages in bilinguals. *Frontiers in Psychology*, *7*(MAY), 1–12. https://doi.org/10.3389/fpsyg.2016.00702

Freeman, M. R., Blumenfeld, H. K., & Marian, V. (2017). Cross-linguistic phonotactic competition and cognitive control in bilinguals. *Journal of Cognitive Psychology*, *29*(7), 783–794. https://doi.org/10.1080/20445911.2017.1321553

Frisch, S. A. (1996). *Similarity and Frequency in Phonology* (Doctoral dissertation). Northwestern University. https://doi.org/https://doi.org/doi:10.7282/T3M61J6F

Frisch, S. A. (2000). Temporally organized lexical representations as phonological units. *Language Acquisition and the Lexicon: Papers in Laboratory Phonology V*, 283–298.

Frisch, S. A. (2004). Language processing and segmental OCP effects. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 346–371). Cambridge University Press. https://doi.org/10.1017/CBO9780511486401.011

Frisch, S. A. (2015). A preliminary investigation of quantitative patterns in sonority sequencing. *Rivista di Linguistica*, *27*(1), 9–27.

Frisch, S. A., & Wright, R. (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics*, *30*(2), 139–162. https://doi.org/10.1006/jpho.2002.0176

Fromkin, V. A. (1971). The Non-Anomalous Nature of Anomalous Utterances. *Language*, *47*(1), 27–52.

Fromkin, V. A. (1973). Slips of the Tongue. *Scientific American*, *229*(6), 110–117. https://doi.org/10.3109/13682826809011435

Fudge, E. C. (1969). Syllables. *Journal of Linguistics*, *5*(2), 253–286.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*(3), 474–496. https://doi.org/10.1353/lan.0.0035

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125. https://doi.org/10.1037/0096-1523.6.1.110

García-Albea, J. E., del Viso, S., & Igoa, J. M. (1989). Movement errors and levels of processing in sentence production. *Journal of Psycholinguistic Research*, *18*(1), 145–161. https://doi.org/10.1007/BF01069052

Garnham, A., Shillcock, R. C., Brown, G. D., Mill, A. I., & Cutler, A. (1982). Slips of the tongue in the London-Lund corpus of spontaneous conversation. *Slips of the Tongue and Language Production*, 251–264. https://doi.org/10.1515/9783110828306.251

Garrett, M. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production vol 1: Speech and talk* (pp. 177–220). Academic Press.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology-Learning Memory and Cognition*, *25*(1), 84–95.

Gaygen, D. E., & Luce, P. A. (1998). Effects of modality on subjective frequency estimates and processing of spoken and printed words.

Perception and Psychophysics, *60*(3), 465–483. https://doi.org/10.3758/BF03206867

Gernsbacher, M. A. (1984). Resolving 20 Years of Inconsistent Interactions Between Lexical Familiarity and Orthography, Concreteness, and Polysemy. *Journal of Experimental Psychology: General*, *113*(2), 256–281. https://doi.org/10.1016/j.micinf.2011.07.011.Innate

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, *66*(3), 363–376. https://doi.org/10.3758/BF03194885

Goad, H. (2011). The Representation of sC Clusters. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The blackwell companion to phonology volume ii. suprasegmental and prosodic phonology* (pp. 1–26). Wiley-Blackwell. https://doi.org/10.1002/9781444335262.wbctp0038

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

Goldrick, M. (2002). *Pattern of sound, patterns in mind: Phonological regularities in speech production* (Doctoral dissertation). Johns Hopkins University. https://doi.org/10.1017/CBO9781107415324.004

Goldrick, M. (2003). Markedness and Frequency in Phonotactic Processing Constraints. *Linguistic Society of America Annual Meeting*, 1–12.

Goldrick, M. (2004). Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language*, *51*(4), 586–603. https://doi.org/10.1016/j.jml.2004.07.004

Goldrick, M., & Blumstein, S. E. (2006a). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*(September 2014), 649–683. https://doi.org/10.1080/01690960500181332

Goldrick, M., & Blumstein, S. E. (2006b). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*(6), 649–683. https://doi.org/10.1080/01690960500181332

Goldrick, M., & Larson, M. (2008). Phonotactic probability influences speech production. *Cognition*, *107*(3), 1155–1164. https://doi.org/10.1016/j.cognition.2007.11.009

Goldstein, R., & Vitevitch, M. S. (2014). The influence of clustering coefficient on word-learning: How groups of similar sounding words facilitate acquisition. *Frontiers in Psychology*, *5*(NOV), 2009–2014. https://doi.org/10.3389/fpsyg.2014.01307

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(3), 515–531. https://doi.org/10.1017/S0305000907008641

Gow, D. W., Melvoid, J., & Manuel, S. (1996). How word onsets drive lexical access and segmentation: evidence from acoustics, phonology and processing. *Spoken Languages*, *1*(1), 66–69. https://doi.org/10.1109/ICSLP.1996.607031

Greenberg, J. H. (1965). Some Generalizations Concerning Initial and Final Consonant Sequences. *Linguistics*, *3*(18), 5–34. https://doi.org/10.1515/ling.1965.3.18.5

Gregory, M. L., Raymond, W. D., Bell, A., Fosler-lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistics Society*, (February), 151–166.

Griffen, T. D. (1981). German affricates. *Lingua*, *53*, 175–198.

Griffin, Z. M., & Ferreira, V. S. (2006). Properties of Spoken Language Production. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 21–60). Academic Press.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & psychophysics*, *28*(4), 267–283. https://doi.org/10.3758/BF03204386

Grosjean, F. (2001). The bilingual ' s language modes. In J. L. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 1–22). Blackwell.

Grossberg, S. (1976). Adaptive Pattern Classification and Universal Recoding: II. Feedback, Expectation, Olfaction, Illusions. *Biological Cybernetics*, *23*, 187–202.

Grossberg, S. (2003). Resonant neural dynamics of speech perception. *Journal of Phonetics*, *31*(3-4), 423–445. https://doi.org/10.1016/S0095-4470(03)00051-2

Grossberg, S., Boardman, I., & Cohen, M. M. (1997). Neural Dynamics of Variable-Rate Speech Categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(2), 481–503. https://doi.org/10.1037/0096-1523.23.2.481

Grossberg, S., & Kazerounian, S. (2011). Laminar cortical dynamics of conscious speech perception: Neural model of phonemic restoration using subsequent context in noise. *The Journal of the Acoustical Society of America*, *130*(1), 440–460. https://doi.org/10.1121/1.3589258

Grossberg, S., & Myers, C. W. (2000). The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. *Psychological Review*, *107*(4), 735–767. https://doi.org/10.1037/0033-295X.107.4.735

Hallé, P. A., & Best, C. T. (2007). Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/ clusters. *The Journal of the Acoustical Society of America*, *121*, 2899–2914. https://doi.org/https://doi.org/10.1121/1.2534656

Hallé, P. A., Segui, J., Frauenfelder, U. H., & Meunier, C. (1998). Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 592–608. https://doi.org/10.1037/0096-1523.24.2.592

Hamza, Y., Okalidou, A., Kyriafinis, G., & Wieringen, A. V. (2018). Sonority's effect as a surface cue on lexical speech perception of children with cochlear implants. *Ear and Hearing*, *39*(5), 992–1007.

Hanulíková, A., McQueen, J. M., & Mitterer, H. (2010). Possible words and fixed stress in the segmentation of Slovak speech. *Quarterly Journal of Experimental Psychology*, *63*(3), 555–79. https://doi.org/10.1080/17470210903038958

Hanulíková, A., Mitterer, H., & McQueen, J. M. (2011). Effects of first and second language on segmentation of non-native speech. *Bilingualism: Language and Cognition*, *14*(4), 506–521. https://doi.org/10.1017/S1366728910000428

Harley, T. A., & Bown, H. E. (1998). What causes a tip-of-the-tongue state? Evidence for lexical neighbourhood effects in speech production. *British Journal of Psychology*, *89*(1), 151–174. https://doi.org/10.1111/j.2044-8295.1998.tb02677.x

Harris, H. D. (2002). Holographic Reduced Representations for Oscillator Recall: A Model of Phonological Production. In W. Gray & C. Schunn (Eds.), *Proceedings of the 24th annual meeting of the cognitive science society* (pp. 423–428). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781315782379-109

Harris, J. (1983). *Syllable structure and stress in Spanish: A nonlinear analysis*. MIT Press.

Harris, J. (1994). *English sound structure*. Blackwell.

Hawkins, S. (2010). Phonological features, auditory objects, and illusions. *Journal of Phonetics*, *38*(1), 60–89. https://doi.org/10.1016/j.wocn.2009.02.001

Hay, J., Pierrehumbert, J. B., & Beckman, M. E. (2004). Speech perception, well-formedness and the statistics of the lexicon. In J. Local, R. Ogden, & R. A. M. Temple (Eds.), *Papers in laboratory phonology vi* (pp. 58–74). Cambridge University Press.

Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactic learning. *Linguistic Inquiry*, *39*(3), 379–440. https://doi.org/10.1162/ling.2008.39.3.379

Hebb, D. O. (1949). *The organization of behavior: a neuropsychological theory*. Wiley.

Heffner, R.-M. S. (1969). *General phonetics* (Reprint). University of Wisconsin.

Heinrich, A., Schneider, B. A., Heinrich, A., & Schneider, B. A. (2010). Elucidating the effects of ageing on remembering perceptually distorted word pairs. *The Quarterly Journal of Experimental Psychology*, *64*(1). https://doi.org/10.1080/17470218.2010.492621

Henke, E., Kaisse, E. M., & Wright, R. (2012). Is the Sonority Sequencing Principle an epiphenomenon? In S. Parker (Ed.), *The sonority controversy* (pp. 65–99). De Gruyter Mouton.

Hofmann, M. J., Stenneken, P., Conrad, M., & Jacobs, A. M. (2007). Sublexical frequency measures for orthographic and phonological units in German. *Behavior Research Methods*, *39*(3), 620–629. https://doi.org/10.3758/BF03193034

Holyoak, K. J., Gentner, D., & Kokinov, B. N. (2001). Introduction: The place of analogy in cognition. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 1–19). MIT Press.

Hoole, P., Bombien, L., Kühnert, B., & Mooshammer, C. (2009). Intrinsic and prosodic effects on articulatory coordination in initial consonant clusters. In G. Fant, H. Fujisaki, & J. Shen (Eds.), *Frontiers in phonetics and speech science* (pp. 275–287). The Commercial Press.

Howell, P., & Rosen, S. (1983). Production and perception of rise time in the voiceless affricate/fricative distinction. *The Journal of the Acoustical Society of America*, *73*(3), 976–984.

Howes, D. (1957). On the Relation between the Intelligibility and Frequency of Occurrence of English Words. *The Journal of the Acous-*

tical Society of America, *29*(2), 296–305. https://doi.org/10.1121/1.1908862

Hume, E. (2003). Language specific markedness: The case of place of articulation. *Studies in Phonetics, Phonology, and Morphology*, *9*, 295–310.

Hume, E., Johnson, K., Seo, M., & Tserdanelis, G. (1999). A cross-linguistic study of stop place perception. *14th International Congress of Phonetic Sciences (ICPhS XIV)*, 2069–2072.

Humphreys, K. R., Menzies, H., & Lake, J. K. (2010). Repeated speech errors: Evidence for learning. *Cognition*, *117*(2), 151–165. https://doi.org/10.1016/j.cognition.2010.08.006

Ibbotson, P. (2013). The scope of usage-based theory. *Frontiers in Psychology*, *4*(MAY), 1–15. https://doi.org/10.3389/fpsyg.2013.00255

Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., & Dupoux, E. (2003). Phonological grammar shapes the auditory cortex : a Functional Magnetic Resonance Imaging Study. *The Journal of Neurosciencce*, *23*(29), 9541–9546.

Jakob, H. (2018). *Produktion von Konsonantenclustern bei Patienten mit aphasisch-phonologischer Störung und Sprechapraxie* (Doctoral dissertation). Ludwig-Maximilians-Universität München.

Jakobson, R. (1962). *Selected writings I: phonological studies*. Mouton.

Janda, R. D., & Joseph, B. D. (2003). Reconsidering the Canons of Sound-Change: Towards a "Big Bang" Theory. In B. J. Blake, K. Burridge, & J. Taylor (Eds.), *Historical linguistics 2001: Selected papers from the 15th international conference on historical linguistics, melbourne, 13-17 august 2001* (pp. 205–219). John Benjamins. https://www.asc.ohio-state.edu/joseph.1/publications/2003bigbang.pdf

Janse, E., & Newman, R. S. (2012). Identifying nonwords: Effects of lexical neighborhoods, phonotactic probability, and listener characteristics. *Language and Speech*, *56*(4), 421–441. https://doi.org/10.1177/0023830912447914

Jarosz, G. (2010). Implicational markedness and frequency in constraint-based computational models of phonological learning. *Journal of Child Language*, *37*(3), 565–606. https://doi.org/10.1017/S0305000910000103

Jun, J. (1995). *Perceptual and Articulatory Factors in Place Assimilation: An Optimality Theoretic Approach* (Doctoral dissertation December). University of California, Los Angeles.

Jurafsky, D. (2003). Probabilistic Modeling in Psycholinguistics: Linguistic Comprehension and Production. In R. Bod, J. Hay, & Jannedy (Eds.), *Probabilistic linguistics* (pp. 39–95). MIT Press.

Jurafsky, D., Bell, A., Gregory, M. L., & Raymond, W. D. (2000). Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. *Frequency and the emergence of linguistic structure*, 229–254.

Jürgens, T., Brand, T., & Kollmeier, B. (2007). Modelling the human-machine gap in speech reception : microscopic speech intelligibility prediction for normal-hearing subjects with an auditory model. *Interspeech*, 410–413.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants′ sensitivity to the sound patterns of native language words. https://doi.org/10.1006/jmla.1993.1022

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants′ sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, *33*(5), 630–645. https://doi.org/10.1006/jmla.1994.1030

Kabak, B., & Idsardi, W. J. (2007). Perceptual distortions in the adaptation of English consonant clusters: syllable structure or consonantal contact constraints? *Language and Speech*, *50*(1), 23–52. https://doi.org/10.1177/00238309070500010201

Kabak, B., Maniwa, K., & Kazanina, N. (2010). Listeners use vowel harmony and word-final stress to spot nonsense words: A study of

Turkish and French. *Laboratory Phonology*, *1*(1), 207–224. https://doi.org/10.1515/labphon.2010.010

Kapatsinski, V. (2014). What is grammar like ? A usage-based constructionist perspective. *Linguistic Issues in Language Technology*, *11*(1), 1–41.

Kawamoto, A. H., & Kello, C. T. (1999). Effect of onset cluster complexity in speeded naming: A test of rule-based approaches. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(2), 361–375. https://doi.org/10.1037/0096-1523.25.2.361

Kazanina, N., Bowers, J. S., & Idsardi, W. J. (2017). Phonemes: Lexical access and beyond. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-017-1362-0

Keller, E. (1987). The Cortical Representation of Motor Processes of Speech. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processes of language* (pp. 125–162). Laurance Erlbaum Associates.

Kittredge, A. K., & Dell, G. S. (2016). Learning to speak by listening : Transfer of phonotactics from perception to production. *Journal of Memory and Language*, *89*, 8–22. https://doi.org/10.1016/j.jml.2015.08.001

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. http://www.cs.indiana.edu/%7B~%7Dport/HDphonol/Klatt.sp.percptn.JPhon.1979.pdf

Klatt, D. H. (1981). Lexical Representations for Speech Production and Perception. In T. Myers, J. D. M. Laver, & J. Anderson (Eds.), *The cognitive representation of speech* (pp. 11–31). North-Holland.

Kohler, K. J. (1977). *Einführung in die Phonetik des Deutschen* (First Edit). Erich Schmidt Verlag.

Kohn, S. E., Melvold, J., & Shipper, V. (1998). The preservation of sonority in the context of impaired lexical-phonological output. *Aphasiology*, *12*(4-5), 375–398. https://doi.org/10.1080/02687039808249539

König, E., & Gast, V. (2012). *Understanding English-German Contrasts* (Third Edit). Erich Schmidt Verlag.

Korecky-Kröll, K., Dressler, W. U., Freiberger, E. M., Reinisch, E., Mörth, K., & Libben, G. (2014). Morphonotactic and phonotactic processing in German-speaking adults. *Language Sciences*, *46*(PA), 48–58. https://doi.org/10.1016/j.langsci.2014.06.006

Krakow, R. A. (1999). Physiological organization of syllables: A review. *Journal of Phonetics*, *27*(1), 23–54. https://doi.org/10.1006/jpho.1999.0089

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. https://doi.org/10.1126/science.1736364

Kwon, H., & Chitoran, I. (2019). Perception of native consonant clusters with non-native phonetic patterns. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th international congress of phonetic sciences* (pp. 388–392).

Ladefoged, P. (1975). *A Course in Phonetics*. Harcourt Brace Jovanovich Inc.

Ladefoged, P. (1997). Linguistic phonetic descriptions. In W. J. Hardcastle & J. D. M. Laver (Eds.), *The handbook of phonetic sciences* (First edit, pp. 589–618). Blackwell.

Laeufer, C. (1995). Effects of tempo and stress on German syllable structure. *Journal of Linguistics*, *31*(1995), 227–266.

Laganaro, M. (2005). Syllable frequency effect in speech production: Evidence from aphasia. *Journal of Neurolinguistics*, *18*(3), 221–235. https://doi.org/10.1016/j.jneuroling.2004.12.001

Laganaro, M., & Alario, F. X. (2006). On the locus of the syllable frequency effect in speech production. *Journal of Memory and Language*, *55*(2), 178–196. https://doi.org/10.1016/j.jml.2006.05.001

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 637–676). De Gruyter Mouton.

Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, *38*(1), 44–59. https://doi.org/10.1016/j.wocn.2010.01.002

Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory and Cognition*, *38*(8), 1137–1146. https://doi.org/10.3758/MC.38.8.1137

Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Theoretical prerequisites*. Stanford University Press.

Langacker, R. W. (2000). A Dynamic Usage-Based Model. In M. Barlow & S. Kemmer (Eds.), *Usage based models of speech* (pp. 1–63). CSLI Publications.

Large, N. R., Frisch, S. A., & Pisoni, D. B. (1998). *Perception of Wordlikeness : Effects of Segment Probability and Length on Subjective Ratings and Processing of Nonword Sound Patterns* (tech. rep. No. 22). Speech Research Laboratory, Department of Psychology, Indiana University. Bloomington, Indiana.

Laver, J. D. M. (1980). Slips of the tongue as neuromuscular evidence for a model of speech production. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of frieda goldman-eisler* (pp. 21–26). Mouton.

Lecumberri, M. L. G., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America*, *119*(4), 2445–2454. https://doi.org/10.1121/1.2180210

Lemhöfer, K., & Broersma, M. (2011). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Lemhöfer, K., Koester, D., & Schreuder, R. (2011). When bicycle pump is harder to read than bicycle bell: Effects of parsing cues in first and second language compound reading. *Psychonomic Bulletin & Review*, *18*(2), 364–370. https://doi.org/10.3758/s13423-010-0044-y

Lentz, T. O. (2011). *Phonotactic illegality and probability in speech perception: Evidence from second language listeners* (Doctoral dissertation). Universiteit Utrecht.

Lentz, T. O., & Kager, R. W. J. (2015). Categorical phonotactic knowledge filters second language input, but probabilistic phonotactic knowledge can still be acquired. *Language and Speech*, *58*(3), 387–413. https://doi.org/10.1177/0023830914559572

Lenzing, A. (2015). Exploring regularities and dynamic systems in L2 development. *Language Learning*, *65*(1), 89–122. https://doi.org/10.1111/lang.12092

Leuninger, H. (1993). *Reden ist Schweigen, Silber ist Gold. Gesammelte Versprecher* (2. Auflage). Deutscher Taschenbuch Verlag GmbH & Co. KG.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*(1-3), 1–22. https://doi.org/10.1016/0010-0277(92)90038-J

Levelt, W. J. M. (1993). Timing in speech production with special reference to word form encoding. *Annals of the New York Academy of Sciences*, *682*(1), 283–295. https://doi.org/10.1111/j.1749-6632.1993.tb22976.x

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–38, discussion 38–75. https://doi.org/10.1017/S0140525X99001776

Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*(1-3), 239–269.

Levitt, A. G., & Healy, A. F. (1985). The roles of phoneme frequency, similarity, and availability in the experimental elicitation of speech errors. *Journal of Memory and Language*, *24*(6), 717–733. https://doi.org/10.1016/0749-596X(85)90055-5

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. https://doi.org/10.1037/h0020279

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5).

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revisited. *Cognition*, *21*, 1–36. https://doi.org/10.3758/PBR.15.2.453

Lieberman, P. (1985). On the evolution of human syntactic ability. Its pre-adaptive Bases-Motor control and speech. *Journal of Human Evolution*, *14*(7), 657–668. https://doi.org/10.1016/S0047-2484(85)80074-9

Lindblom, B. (1983). Economy of Speech Gestures. *The production of speech* (pp. 217–245). Springer.

Linzen, T., & Gallagher, G. (2014). The Timecourse of Generalization in Phonotactic Learning. *Proceedings of the Annual Meetings on Phonology*, *1*(1). https://doi.org/10.1378/chest.115.1.144

Lléo, C., & Prinz, M. (1997). Syllable Structure Parameters and the Acquisition of Affricates. In S. Hannahs & M. Young-Scholten (Eds.), *Focus on phonological acquisition* (pp. 143–163).

Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception and Psychophysics*, *62*(3), 615–625. https://doi.org/10.3758/BF03212113

Luce, P. A., Goldinger, S. D., & Vitevitch, M. S. (2000). It's good . . . but is it ART? *Behavioral and Brain Sciences*, *23*(3), 336. https://doi.org/10.1017/S0140525X00343242

Luce, P. A., & Large, N. R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes*, *16*(5-6), 565–581. https://doi.org/10.1080/01690960143000137

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The Neighborhood Activation Model. *Ear and Hearing*, *19*(1), 1–36. https://doi.org/10.1097/00003446-199802000-00001

MacKay, D. G. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology*, *3*(2), 210–227.

MacKay, D. G. (1978). Speech errors inside the syllable. In A. Bell & J. B. Hooper (Eds.), *Syllables and segments* (pp. 201–212). North-Holland.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. *Psychological Review*, *89*(5), 483–506. https://doi.org/10.1037/0033-295X.89.5.483

MacKenzie, H., Curtin, S., & Graham, S. A. (2012). 12-Month-Olds' Phonotactic Knowledge Guides Their Word-Object Mappings. *Child Development*, *83*(4), 1129–1136. https://doi.org/10.1111/j.1467-8624.2012.01764.x

Madlener, K. (2016). Input optimization: Effects of type and token frequency manipulations in instructed second language learning. In H. Behrens & S. Pfänder (Eds.), *Experience counts: Frequency effects in language* (pp. 133–174). Walter de Gruyter GmbH & Co KG.

Marchegiani, L., & Fafoutis, X. (2015). On cross-language consonant identification in second language noise. *The Journal of the Acoustical Society of America*, *138*(4), 2206–2208. https://doi.org/10.1121/1.2816563

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PloS One*, *7*(8), 1–11. https://doi.org/10.1371/journal.pone.0043230

Marian, V., Blumenfeld, H. K., & Boukrina, O. V. (2008). Sensitivity to phonological similarity within and across languages. *Journal of Psycholinguistic Research*, *37*(3), 141–170. https://doi.org/10.1007/s10936-007-9064-9

Marian, V., & Spivey, M. J. (1999). Activation of Russian and English cohorts during bilingual spoken word recognition. In M. Hahn & S. Stoness (Eds.), *Proceedings of the 21st annual conference of the cognitive science society* (pp. 349–354). Erlbaum.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, *189*(4198), 226–228. https://doi.org/10.1126/science.189.4198.226

Martin, A., & Peperkamp, S. (2017). Assessing the distinctiveness of phonological features in word recognition: Prelexical and lexical influences. *Journal of Phonetics*, *62*, 1–11. https://doi.org/10.1016/j.wocn.2017.01.007

Massaro, D. W. (1972). Perceptual images, processing time and perceptional units in auditory perception. *Psychological Review*, *79*(2), 124–145.

Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception. *Perception & Psychophysics*, *34*(4), 338–348. https://doi.org/10.3758/BF03203046

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

McClelland, J. L., & Elman, J. L. (1986). The TRACE Model of Speech. *Cognitive Psychology*, *18*, 1–86.

McMillan, C. T. (2008). *Articulatory Evidence for Interactivity in Speech Production* (Doctoral dissertation). University of Edinburgh. http://hdl.handle.net/1842/3280

McMillan, C. T., & Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition*, *117*(3), 243–260. https://doi.org/10.1016/j.cognition.2010.08.019

McQueen, J. M., & Cutler, A. (2010). Cognitive Processes in Speech Perception. In W. J. Hardcastle, J. Larver, & F. E. Gibbon (Eds.), *The handbook of phonetic sciences* (Second Edi, pp. 489–520). Wiley-Blackwell.

McQueen, J. M., & Cutler, A. (2013). Cognitive processes in speech perception. *The Handbook of Phonetic Sciences*, 489–520.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30*, 1113–1126. https://doi.org/10.1207/s15516709cog0000

McQueen, J. M., & Pitt, M. A. (1996). Transitional probability and phoneme monitoring. *Proceedings of ICSLP 96*, (4), 2502–2505.

Meffert, E., Tillmanns, E., Heim, S., Jung, S., Huber, W., & Grande, M. (2011). Taboo: A Novel Paradigm to Elicit Aphasia-Like Trouble-Indicating Behaviour in Normally Speaking Individuals. *Journal of Psycholinguistic Research*, *40*(5), 307–326. https://doi.org/10.1007/s10936-011-9170-6

Mehler, J. (1981). The Role of Syllables in Speech Processing : Infant and Adult Data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *295*(1077), 333–352.

Meringer, R., & Mayer, K. (1895). *Versprechen und Verlesen*. Göschen.

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in Human Superior Temporal Gyrus. *Science*, *343*(January), 1–6. https://doi.org/10.1126/science.1245994

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(6), 1146–1160. https://doi.org/10.1037/0278-7393.17.6.1146

Meyer, B. T., Jürgens, T., Wesker, T., Brand, T., & Kollmeier, B. (2010). Human phoneme recognition depending on speech-intrinsic variability. *The Journal of the Acoustical Society of America*, *128*(5), 3126–41. https://doi.org/10.1121/1.3493450

Miller, R. T. (2011). Impact of L2 Reading Proficiency on L1 Transfer in Visual Word Recognition. In G. Granena, J. Koeth, A. Lukyanchenko, S. Lee-Ellis, G. P. Botana, & E. Rhodes (Eds.), *Selected proceedings of the 2010 second language research forum* (pp. 78–90). http://www.lingref.com/cpp/slrf/2010/index.html

Miozzo, M., & Buchwald, A. (2013). On the nature of sonority in spoken word production : Evidence from neuropsychology. *Cognition*, *128*(3), 287–301. https://doi.org/10.1016/j.cognition.2013.04.006

Mitani, S., Kitama, T., & Sato, Y. (2006). Voiceless affricate/fricative distinction by frication duration and amplitude rise slope. *The Journal of the Acoustical Society of America*, *120*(3), 1600–1607.

Mitterer, H., Scharenborg, O., & McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition*, *129*(2), 356–361. https://doi.org/10.1016/j.cognition.2013.07.011

Moerk, E. L. (1980). Relationships between parental input frequencies and children's language acquisition: A reanalysis of Brown's data. *Journal of Child Language*, *7*(1), 105–118. https://doi.org/10.1017/S0305000900007054

Mooshammer, C., Tiede, M., Katsika, A., & Goldstein, L. (2015). Effects of phonological competition on speech planning and execution. In T. S. C. f. I. 2015 (Ed.), *Proceedings of the 18th international congress of phonetic sciences* (pp. 6–10). University of Glasgow. http://www.icphs2015.info/pdfs/Papers/ICPHS1041.pdf

Morais, J. (2021). The phoneme: A conceptual heritage from alphabetic literacy. *Cognition*, *213*(August 2020). https://doi.org/10.1016/j.cognition.2021.104740

Morelli, F. (1999). *The phonotactics and phonology of obstruent clusters in Optimality Theory* (Doctoral dissertation). University of Maryland.

Moreno-Torres, I., Otero, P., Luna-Ramírez, S., & Garayzábal Heinze, E. (2017). Analysis of Spanish consonant recognition in 8-talker babble. *The Journal of the Acoustical Society of America*, *141*(5), 3079–3090. https://doi.org/10.1121/1.4982251

Moreton, E. (2002). Structural constraints in the perception of English stop-sonorant clusters. *Cognition*, *84*(1), 55–71. https://doi.org/10.1016/S0010-0277(02)00014-8

Motley, M. T., & Baars, B. J. (1975). Encoding Sensitivities To Phonological Markedness and Transitional Probability: Evidence From Spoonerisms. *Human Communication Research*, *1*(4), 353–361. https://doi.org/10.1111/j.1468-2958.1975.tb00284.x

Munson, B. (2001). Phonological Pattern Frequency and Speech Production in Children and Adults. *Journal of Speech, Language, and Hearing Research*, *44*(4), 778–792.

Nathan, G. S. (1996). Steps towards a cognitive phonology. In B. Hurch & R. A. Rhodes (Eds.), *Natural phonology: The state of the art* (pp. 107–120). Mouton de Gruyter.

Newton, C. N. (1972). *Perceptual confusions among permissible and impermissible English consonant clusters* (Master's Thesis). University of British Columbia.

Nooteboom, S. G. (1973). The Tongue Slips into Patterns. In V. A. Fromkin (Ed.), *Speech errors as linguistic evidence* (pp. 144–156). Mouton.

Nooteboom, S. G. (2005). Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. *Speech Communication*, *47*(1-2), 43–58. https://doi.org/10.1016/j.specom.2005.02.003

Nooteboom, S. G., & Quené, H. (2015). The word onset effect: Some contradictory findings. In R. Eklund (Ed.), *Proceedings of diss 2015, the 7th workshop on disfluencies in spontaneous speech, 8-9 august 2015* (pp. 69–72).

Ogden, R. (2009). *An Introduction to English Phonetics*. Edinburgh University Press.

Ohala, D. K. (1999). The influence of sonority on children's cluster reductions. *Journal of Communication Disorders*, *32*(6), 397–422.

Ohala, J. J. (1992). Alternatives to the Sonority Hierarchy for Explaining Segmental Sequential Constraints. *Papers from the Parasession on the Syllable*, 319–338. papers2://publication/uuid/A86D412B-B282-431C-92E1-293C90697FA0

Ohala, J. J., & Kawasaki, H. (1984). Prosodic phonology and phonetics. *Phonology*, *1*(1984), 113–127. https://doi.org/10.1017/S0952675700000312

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*(4), 273–281. https://doi.org/10.1080/17470216508416445

Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition*, *83*(1), 13–23. https://doi.org/10.1016/S0010-0277(01)00165-2

Orzechowska, P., & Wiese, R. (2015). Preferences and variation in word-initial phonotactics : A multi-dimensional evaluation of German and Polish. *Folia Linguistica*, *49*(2), 439–486. https://doi.org/10.1515/flin-2015-0016

O'Seaghdha, P. G., Chen, J. Y., & Chen, T. M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition*, *115*(2), 282–302. https://doi.org/10.1016/j.cognition.2010.01.001

Osterhout, L., & Mobley, L. A. (1995). Event-Related Brain Potentials elicited by failure to agree. https://doi.org/10.1006/jmla.1995.1033

Ott, S., van de Vijver, R., & Höhle, B. (2006). The effect of phonotactic constraints in German-speaking children with delayed phonological acquisition: Evidence from production of word-initial consonant clusters. *Advances in Speech-Language Pathology*, *8*(December), 323–334. https://doi.org/10.1080/14417040600970622

Page, M. P., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3737–3753. https://doi.org/10.1098/rstb.2009.0173

Parker, S. (2002). *Quantifying the Sonority Hierarchy* (Doctoral dissertation). University of Massachusetts Amherst. https://doi.org/10.1007/s13398-014-0173-7.2

Parker, S. (2017). Sounding out Sonority. *Linguistics and Language Compass*, *11*(9), 1–197. https://doi.org/10.1111/lnc3.12248

Perret, C., Schneider, L., Dayer, G., & Laganaro, M. (2014). Convergences and divergences between neurolinguistic and psycholinguistic data in the study of phonological and phonetic encoding: A parallel investigation of syllable frequency effects in braindamaged and healthy speakers. *Language, Cognition and Neuroscience*, *29*(6), 714–727. https://doi.org/10.1080/01690965.2012.678368

Perruchet, P., & Peereman, R. (2004). The exploitation of distributional information in syllable processing. *Journal of Neurolinguistics*, *17*(2-3), 97–119. https://doi.org/10.1016/S0911-6044(03)00059-9

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137–158). John Benjamins Publishing.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*(2-3), 115–154.

Pisoni, D. B., & Sawusch, J. R. (1975). Some Stages of Processing in Speech Perception. In A. Cohen & S. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 16–35). Springer-Verlag.

Pitt, M. A. (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception & Psychophysics*, *60*(6), 941–951. https://doi.org/10.3758/BF03211930

Pitt, M. A., & McQueen, J. M. (1998). Is Compensation for Coarticulation Mediated by the Lexicon? *Journal of Memory and Language*, *39*, 347–370. https://doi.org/10.1006/jmla.1998.2571

Pitt, M. A., Myung, J. I., & Altteri, N. (2007). Modeling the word recognition data of Vitevitch and Luce (1998): Is it ARTful? *Psychonomic Bulletin and Review*, *14*(3), 442–448. https://doi.org/10.3758/BF03194086

Pouplier, M. (2008). The role of a coda consonant as error trigger in repetition tasks. *Journal of Phonetics*, *36*(1), 114–140. https://doi.org/10.1016/j.wocn.2007.01.002

Pouplier, M. (2012). The gestural approach to syllable structure: Universal, language- and cluster-specific aspects. In S. Fuchs, M. Weirich, D. Pape, & P. Perrier (Eds.), *Speech planning and dynamics* (pp. 63–96). Peter Lang.

Pouplier, M., & Goldstein, L. (2010). Intention in articulation: Articulatory timing in alternating consonant sequences and its implications for models of speech production. *Language and Cognitive Processes*, *25*(5), 616–649. https://doi.org/10.1080/01690960903395380

Pouplier, M., Marin, S., Hoole, P., & Kochetov, A. (2017). Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics*, *64*, 108–126. https://doi.org/10.1016/j.wocn.2017.01.006

Pouplier, M., Marin, S., & Kochetov, A. (2017). The difficulty of articulatory complexity. *Cognitive Neuropsychology*, *34*(7-8), 472–475. https://doi.org/10.1080/02643294.2017.1419947

Pouplier, M., Marin, S., & Waltl, S. (2014). Voice Onset Time in Consonant Cluster Errors: Can Phonetic Accommodation Differentiate Cognitive from Motor Errors? *Journal of Speech, Language, and Hearing Research*, *57*(October), 1577–1588. https://doi.org/10.1044/2014

Pylkkänen, L., Stringfellow, A., & Marantz, A. (2002). Neuromagnetic evidence for the timing of lexical activation: An MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain and Language*, *81*(1-3), 666–678. https://doi.org/10.1006/brln.2001.2555

R Core Team. (2016). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.r-project.org/

Raettig, T., & Kotz, S. A. (2008). Auditory processing of different types of pseudo-words: An event-related fMRI study. *NeuroImage*, *39*, 1420–1428. https://doi.org/10.1016/j.neuroimage.2007.09.030

Reason, J. T. (1992). Cognitive Underspecification. Its Variety and Consequences. In B. J. Baars (Ed.), *Experimental slips and human error: Exploring the architecture of volition* (pp. 71–91). Plenum Press.

Reetz, H., & Jongman, A. (2009). *Phonetics. Transcription, Production, Acoustics, and Perception*. Wiley-Blackwell.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (pp. 243–335). Academic Press.

Riecker, A., Brendel, B., Ziegler, W., Erb, M., & Ackermann, H. (2008). The influence of syllable onset complexity and syllable frequency on speech motor control. *Brain and Language*, *107*(2), 102–113. https://doi.org/10.1016/j.bandl.2008.01.008

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*(3), 249–284. https://doi.org/10.1016/S0010-0277(97)00027-9

Rohe, U. (2019). *The progressive in present-day spoken English Real-time studies of its spread and functional diversification* (Doctoral dissertation). Albert-Ludwigs-Universität Freiburg.

Romani, C., & Calabrese, A. (1998). Syllabic constraints in the phonological errors of an aphasic patient. *Brain and Language*, *64*(1), 83–121. https://doi.org/10.1006/brln.1998.1958

Romani, C., & Galluzzi, C. (2005). *Effects of syllabic complexity in predicting accuracy of repetition and direction of errors in patients with articulatory and phonological difficulties* (Vol. 22). https://doi.org/10.1080/02643290442000365

Romani, C., Galluzzi, C., Bureca, I., & Olson, A. (2011). Effects of syllable structure in aphasic errors: Implications for a new model of speech production. *Cognitive Psychology*, *62*(2), 151–192. https://doi.org/10.1016/j.cogpsych.2010.08.001

Romani, C., Galluzzi, C., & Goslin, J. (2016). Phoneme and syllable frequency effects in the errors of aphasic patients: Syllables are structures not 'chunks'. *Paper presented at the 54th Annual Academy of Aphasia Meeting*. https://doi.org/10.3389/conf.fpsyg.2016.68.00111

Romani, C., Galluzzi, C., Goslin, J., Bureca, I., & Olson, A. (2013). Sonority, Frequency and Markedness in Errors of Aphasic Patients. *Procedia - Social and Behavioral Sciences*, *94*(0), 55–56. https://doi.org/http://dx.doi.org/10.1016/j.sbspro.2013.09.024

Romani, C., Galluzzi, C., Guariglia, C., & Goslin, J. (2017). Comparing phoneme frequency, age of acquisition, and loss in aphasia: Implications for phonological universals. *Cognitive Neuropsychology*, *34*(7-8), 449–471. https://doi.org/10.1080/02643294.2017.1369942

Rossi, S., Jürgenson, I. B., Hanulíková, A., Telkemeyer, S., Wartenburger, I., & Obrig, H. (2011). Implicit processing of phonotactic cues: Evidence from electrophysiological and vascular responses. *Journal of Cognitive Neuroscience*, *23*(7), 1752–1764. https://doi.org/10.1162/jocn.2010.21547

Rutter, B. (2011). Acoustic analysis of a sound change in progress: The consonant cluster /s/ in English. *Journal of the International Phonetic Association*, *41*(1), 27–40. https://doi.org/10.1017/S0025100310000307

Sadat, J., Martin, C. D., Costa, A., & Alario, F. X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology*, *68*, 33–58. https://doi.org/10.1016/j.cogpsych.2013.10.001

Samuel, A. G. (1989). Insights from a failure of selective adaptation: Syllable-initial and syllable-final consonants are different. *Perception & Psychophysics*, *45*(6), 485–493. https://doi.org/10.3758/BF03208055

Santiago, J., Pérez, E., Palma, A., & Stemberger, J. P. (2007). Syllable, Word, and Phoneme Frequency Effects in Spanish phonological speech errors: The David effect on the source of the error. *MIT Working Papers in Linguistics*, *53*, 265–303.

Savin, H. B. (1963). Word-Frequency Effect and Errors in the Perception of Speech. *The Journal of the Acoustical Society of America*, *35*(2), 200–206.

Scherer, G., & Wollmann, A. (1986). *Englische Phonetik und Phonologie* (Third Edit). Erich Schmidt Verlag.

Schneider, U. (2014). *Frequency, Chunks & Hesitations: A Usage-based Analysis of Chunking in English* (Doctoral dissertation). Albert-Ludwigs-Universität Freiburg, NIHIN studies.

Scholes, R. J. (1966). *Phonotactic grammaticality*. Walter de Gruyter GmbH & Co KG.

Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, *14*(3), 369–385. https://doi.org/10.1017/S0142716400010845

Segawa, J., Masapollo, M., Tong, M., Smith, D. J., & Guenther, F. H. (2019). Chunking of phonological units in speech sequencing. *Brain and Language*, *195*(May), 104636. https://doi.org/10.1016/j.bandl.2019.05.001

Seibert Hanson, A. E., & Carlson, M. T. (2014). The Roles of First Language and Proficiency in L2 Processing of Spanish Clitics: Global Effects. *Language Learning*, *64*(2), 310–342. https://doi.org/10.1111/lang.12050

Selkirk, E. O. (1982). The syllable. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations (part ii)* (pp. 337–383). Foris Publications.

Selkirk, E. O. (1984). *On the major class features and syllable theory*. MIT Press.

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition*, *53*(2), 91–127. https://doi.org/10.1016/0010-0277(94)90067-1

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, *133*(1), 140–155. https://doi.org/10.1016/j.cognition.2014.06.013

Shapiro, M. (1995). A Case of Distant Assimilation : / str / → / ʃtr /. *American Speech*, *70*(1), 101–107. https://www.jstor.org/stable/455876

Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, *18*(1), 41–55. https://doi.org/10.1016/S0022-5371(79)90554-1

Shuster, L. I. (2009). The effect of sublexical and lexical frequency on speech production : An fMRI investigation. *Brain and Language*, *111*(1), 66–72. https://doi.org/10.1016/j.bandl.2009.06.003

Sievers, E. (1897). *Grundzüge der Lautphysiologie zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen*. Breitkopf & Härtel.

Silbert, N. H., & Zadeh, L. M. (2015). Variability in noise-masked consonant identification. *18th International Congress of Phonetic Sciences*.

Slis, A. W., & van Lieshout, P. (2016). The effect of phonetic context on the dynamics of intrusions and reductions. *Journal of Phonetics*, *57*, 1–20. https://doi.org/10.1016/j.wocn.2016.04.001

Song, J., Martin, L., & Iverson, P. (2019). Native and non-native speech recognition on noise : Neural measures of auditory and lexical processing. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *International congress of phonetic sciences* (pp. 5–9).

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: partial activation of an irrelevant lexicon. *Psychological Science*, *10*(3), 281–284. https://doi.org/10.1111/1467-9280.00151

Staiger, A., & Ziegler, W. (2008). Syllable frequency and syllable structure in the spontaneous speech production of patients with apraxia of speech. *Aphasiology*, *22*(11), 1201–1215.

Steinberg, J., Jacobsen, T. K., & Jacobsen, T. (2016). The Development of English as a Second Language With and Without Specific Language Impairment: Clinical Implications. *Journal of Speech, Language, and Hearing Research*, *59*, 557–571. https://doi.org/10.1044/2015

Steinberg, J., Truckenbrodt, H., & Jacobsen, T. (2011). Phonotactic constraint violations in German grammar are detected automatically in auditory speech processing: A human event-related potentials study. *Psychophysiology*, *48*(9), 1208–1216. https://doi.org/10.1111/j.1469-8986.2011.01200.x

Stemberger, J. P. (1991). Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language*, *30*, 151–185. https://doi.org/10.1016/0749-596x(91)90002-2

Stemberger, J. P. (1993). Spontaneous and Evoked Slips of the Tongue. In G. Blanken, J. Dittmann, H. Grimm, J. C. Marshall, & C.-W. Wallesch (Eds.), *Linguistic disorders and pathologies: An international handbook* (pp. 53–65). de Gruyter.

Stemberger, J. P. (2004). Neighbourhood effects on error rates in speech production. *Brain and Language*, *90*(1-3), 413–422. https://doi.org/10.1016/S0093-934X(03)00452-8

Stemberger, J. P., & Treiman, R. (1986). The internal structure of word-initial consonant clusters. *Journal of Memory and Language*, *25*(2), 163–180. https://doi.org/10.1016/0749-596X(86)90027-6

Stenneken, P., Bastiaanse, R., Huber, W., & Jacobs, A. M. (2005). Syllable structure and sonority in language inventory and aphasic neologisms. *Brain and Language*, *95*(2), 280–292. https://doi.org/10.1016/j.bandl.2005.01.013

Stevens, M., & Harrington, J. (2016). The phonetic origins of /s/-retraction: Acoustic and perceptual evidence from Australian English. *Journal of Phonetics*, *58*, 118–134. https://doi.org/10.1016/j.wocn.2016.08.003

Stevens, M., & Loakes, D. (2019). Individual differences and sound change actuation : evidence from imitation and perception of English /str/. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th international congress of phonetic sciences* (pp. 3200–3204).

Stites, J., Demuth, K., & Kirk, C. (2004). Markedness vs. frequency effects in coda acquisition. *Proceedings of the 28th annual Boston University conference on language development*, (9870676), 565–576.

Studdert-Kennedy, M. (1998). The particulate origins of language generativity: from syllable to gesture. In J. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language* (pp. 202–221). Cambridge University Press.

Sussman, H. M. (1984). A neuronal model for syllable representation. *Brain and Language*, *22*(1), 167–177. https://doi.org/10.1016/0093-934X(84)90087-7

Tajima, K., Erickson, D., & Nagao, K. (2003). *Production of syllable structure in a second language: Factors affecting vowel epenthesis in Japanese-accented English* (No. *2*), Indiana University.

Tamási, K., & Berent, I. (2015). Sensitivity to Phonological Universals: The Case of Stops and Fricatives. *Journal of Psycholinguistic Research*, *44*(4), 359–381. https://doi.org/10.1007/s10936-014-9289-3

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Taylor, R. (1990). Interpretation of the correlation coefficient: A basic review. *Journal of Diagnostic Medical Sonography*, *6*(1), 35–39.

Tilsen, S. (2011). Metrical regularity facilitates speech planning and production. *Laboratory Phonology*, *2*(1), 185–218. https://doi.org/10.1515/labphon.2011.006

Tilsen, S. (2016). Selection and coordination: The articulatory basis for the emergence of phonological structure. *Journal of Phonetics*, *55*, 53–77. https://doi.org/10.1016/j.wocn.2015.11.005

Trapman, M., & Kager, R. W. J. (2009). The Acquisition of Subset and Superset Phonotactic Knowledge in a Second Language. *Language Acquisition*, *16*(3), 178–221. https://doi.org/10.1080/10489220903011636

Treiman, R. (1986). The Division between Onsets and Rimes in English. *Journal of Memory and Language*, *25*(4), 476–491. https://doi.org/10.1016/0749-596X(86)90039-2

Treiman, R., Salasoo, A., Slowiaczek, L. M., & Pisoni, D. B. (1982). *Effects of syllable structure on adults' phoneme monitoring performance* (tech. rep. Progress Report No. 8). Speech Research Laboratory, Indiana University. Indiana.

Tremblay, P., Deschamps, I., Baroni, M., & Hasson, U. (2016). Neural sensitivity to syllable frequency and mutual information in speech perception and production. *NeuroImage*, *136*, 106–121. https://doi.org/10.1016/j.neuroimage.2016.05.018

Trubetzkoy, N. S. (1939). *Grundzüge der Phonologie [Foundations of phonology]*. Traveaux du Cercle Linguistique de Prague.

Tzakosta, M. (2009). Asymmetries in /s/ cluster production and their implications for language learning and language teaching. *aper presented at the 18th Iinternational Symposium of Theoretical and Applied Linguistics*, (2000), 335–373.

Ulbrich, C., Alday, P. M., Knaus, J., Orzechowska, P., & Wiese, R. (2016). The role of phonotactic principles in language processing. *Lan-*

guage, Cognition and Neuroscience, *31*(5), 662–682. https://doi.org/10.1080/23273798.2015.1136427

Ulbrich, C., & Wiese, R. (2018). Phonotactic principles and exposure in second language processing. In C. Ulbrich, A. Werth, & R. Wiese (Eds.), *Empirical approaches to the phonological structure of words* (pp. 153–179). de Gruyter. https://doi.org/10.1515/9783110542899-007

van der Lugt, A. H. (2001). The use of sequential probabilities in the segmentation of speech. *Perception and Psychophysics*, *63*(5), 811–823. https://doi.org/10.3758/BF03194440

van de Vijver, R., & Baer-Henney, D. (2012). Sonority intuitions are provided by the lexicon. In S. Parker (Ed.), *The sonority controversy* (pp. 95–98). https://doi.org/10.1017/S002510030800371X

van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, *38*(4), 584–589. https://doi.org/10.3758/BF03193889

van Hessen, A. J., & Schouten, M. E. H. (1999). Categorical perception as a function of stimulus quality. *Phonetica*, *56*(1-2), 56–72. https://doi.org/28441

Vennemann, T. (2011). *Preference laws for syllable structure: And the explanation of sound change with special reference to German, Germanic, Italian, and Latin*. Walter de Gruyter.

Verwey, W. B., & Abrahamse, E. L. (2012). Distinct modes of executing movement sequences: Reacting, associating, and chunking. *Acta Psychologica*, *140*(3), 274–282. https://doi.org/10.1016/j.actpsy.2012.05.007

Vitevitch, M. S. (2002). Naturalistic and Experimental Analyses of Word Frequency and Neighborhood Density Effects in Slips of the Ear. *Language and Speech*, *45*(Pt 4), 407–434. https://doi.org/10.1016/j.biotechadv.2011.08.021.Secreted

Vitevitch, M. S. (2003). The influence of sublexical and lexical representations on the processing of spoken words in English. *Clinical*

Linguistics and Phonetics, *17*(6), 487–499. https://doi.org/https://doi.org/10.1080/0269920031000107541

Vitevitch, M. S., Armbrüster, J., & Chu, S. (2004). Sublexical and Lexical Representations in Speech Production: Effects of Phonotactic Probability and Onset Density. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 514–529. https://doi.org/10.1037/0278-7393.30.2.514

Vitevitch, M. S., & Castro, N. (2015). Using network science in the language sciences and clinic. *International Journal of Speech-Language Pathology*, *17*(1), 1–25. https://doi.org/10.3109/17549507.2014.987819

Vitevitch, M. S., & Luce, P. A. (1998). When Words Compete: Levels of Processing in Perception of Spoken Words. *Psychological Science*, *9*(4), 325–329. https://doi.org/10.1111/1467-9280.00064

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*(3), 374–408. https://doi.org/10.1006/jmla.1998.2618

Vitevitch, M. S., & Luce, P. A. (2005). Increases in phonotactic probability facilitate spoken nonword repetition. *Jounal of Memory and Language*, *52*, 193–204. https://doi.org/10.1016/j.jml.2004.10.003

Vitevitch, M. S., & Luce, P. A. (2016). Phonological Neighborhood Effects in Spoken Word Perception and Production. *Annual Review of Linguistics*, *2*(1), 7.1–7.20. https://doi.org/10.1146/annurev-linguist-030514-124832

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: implications for the processing of spoken nonsense words. *Language and Speech*, *40*(1), 47–62. https://doi.org/10.1177/002383099704000103

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words.

Brain and Language, *68*(1-2), 306–311. https://doi.org/10.1006/brln.1999.2116

Vitevitch, M. S., & Stamer, M. (2006). The curious case of competition in Spanish speech production. *Language and Cognitive Processes*, *21*(6), 760–770. https://doi.org/10.1080/01690960500287196

Vitevitch, M. S., Storkel, H. L., Francisco, A. C., Evans, K. J., & Goldstein, R. (2014). The influence of known-word frequency on the acquisition of new neighbours in adults : evidence for exemplar representations in word learning. *Language, Cognition and Neuroscience*, *29*(10), 1311–1316. https://doi.org/10.1080/23273798.2014.912342

Vousden, J. I., Brown, G. D., & Harley, T. a. (2000). Serial Control of Phonology in Speech Production: A Hierarchical Model. *Cognitive Psychology*, *41*(2), 101–175. https://doi.org/10.1006/cogp.2000.0739

Vousden, J. I., & Maylor, E. (2006). Speech errors across the lifespan. *Language and Cognitive Processes*, *21*(1-3), 48–77. https://doi.org/10.1080/01690960400001838

Wade, T., Dogil, G., Schütze, H., Walsh, M., & Möbius, B. (2010). Syllable frequency effects in a context-sensitive segment production model. *Journal of Phonetics*, *38*, 227–239. https://doi.org/10.1016/j.wocn.2009.10.004

Wagner, M., Shafer, V. L., Martin, B., & Steinschneider, M. (2012). The phonotactic influence on the perception of a consonant cluster /pt/ by native English and native Polish listeners: A behavioral and event related potential (ERP) study. *Brain and Language*, *123*(1), 30–41. https://doi.org/doi:10.1016/j.bandl.2012.06.002

Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning Memory and Cognition*, *32*(2), 387–398. https://doi.org/10.1037/0278-7393.32.2.387

Warker, J. A., Xu, Y., Dell, G. S., & Fisher, C. (2009). Speech errors reflect the phonotactic constraints in recently spoken syllables, but not in recently heard syllables. *Cognition*, *112*(1), 81–96. https://doi.org/10.1016/j.cognition.2009.03.009

Warner, N., Smits, R., McQueen, J. M., & Cutler, A. (2005). Phonological and statistical effects on timing of speech perception: Insights from a database of Dutch diphone perception. *Speech Communication*, *46*(1), 53–72. https://doi.org/10.1016/j.specom.2005.01.003

Warren, R. M. (1984). Perceptual restoration of obliterated sounds. *Psychological Bulletin*, *96*(2), 371–383. https://doi.org/10.1037/0033-2909.96.2.371

Weber, A. (2001). *Language-specific listening: The case of phonetic sequences* (Doctoral dissertation).

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1–25. https://doi.org/10.1016/S0749-596X(03)00105-0

Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *The Journal of the Acoustical Society of America*, *119*(1), 597–607. https://doi.org/10.1121/1.2141003

Wells, R. (1951). Predicting slips of the tongue. *Speech errors as linguistic evidence* (pp. 82–87).

Whitney, W. D. (1874). The relation of vowel and consonant. In W. D. Whitney (Ed.), *Oriental and linguistic studies. second series. the east and west; religion and mythology; orthography and phonology; hindu astronomy* (pp. 277–300). Scribner, Armstrong, Company.

Wickelgren, W. A. (1969). Context-Sensitive Coding in Speech Recognition, Articulation and Development. In K. N. Leibovic (Ed.), *Information processing in the nervous system* (pp. 85–86).

Wiese, R., Orzechowska, P., Alday, P. M., & Ulbrich, C. (2017). Structural principles or frequency of use? An erp experiment on the learnability of consonant clusters. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.02005

Wilshire, C. E. (1998). Serial order in phonological encoding: an exploration of the 'word onset effect' using laboratory-induced errors.

Cognition, *68*(2), 143–66. https://doi.org/S0010-0277(98)00045-6[pii]

Wilshire, C. E. (1999). The “Tongue Twister” Paradigm as a Technique for Studying Phonological Encoding. *Language and Speech*, *42*(1), 57–82. https://doi.org/10.1177/00238309990420010301

Woods, D. L., Yund, E. W., Herron, T. J., & Cruadhlaoich, M. A. I. U. (2010). Consonant identification in consonant-vowel-consonant syllables in speech-spectrum noise. *The Journal of the Acoustical Society of America*, *127*(3), 1609–1623. https://doi.org/10.1121/1.3293005

Wright, R. (2001). Perceptual Cues in Contrast Maintenance. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 251–277). Academic Press.

Wright, R. (2004). A review of perceptual cues and cue robustness. https://doi.org/10.1017/CBO9780511486401.002

Yang, C. (2015). For and against frequencies. *Journal of Child Language*, *42*(2), 287–293. https://doi.org/10.1017/S0305000914000683

Yavaş, M. (2003). Role of sonority in developing phonologies. *Journal of Multilingual Communication Disorders*, *1*(2), 79–98. https://doi.org/10.1080/1476967021000048140

Yazawa, K., Konishi, T., Hanzawa, K., Short, G., & Kondo, M. (2015). Vowel epenthesis in Japanese speakers’ L2 English. In The Scottish Consortium for ICPhS2015 (Ed.), *Proceedings of the 18th international congress of phonetic sciences* (969:1–4). University of Glasgow.

Yip, M. (2000). Recognition Of Spoken Words in the Continuous Speech: Effects of Transitional Probability. *Proceedings of the Sixth International Conference on Spoken Language Processing (ICSLP 2000)*, 758–761.

Zamuner, T. S., Morin-Lessard, E., Strahm, S., & Page, M. P. (2016). Spoken word recognition of novel words, either produced or only heard during learning. *Journal of Memory and Language*, *89*, 55–67. https://doi.org/10.1016/j.jml.2015.10.003

Zhao, X., & Berent, I. (2016). Universal Restrictions on Syllable Structure: Evidence From Mandarin Chinese. *Journal of Psycholinguistic Research*, *45*(4), 795–811. https://doi.org/10.1007/s10936-015-9375-1

Ziegler, W., & Aichert, I. (2015). How much is a word? Predicting ease of articulation planning from apraxic speech error patterns. *Cortex*, *69*, 24–39. https://doi.org/10.1016/j.cortex.2015.04.001

Ziegler, W., Thelen, A. K., Staiger, A., & Liepold, M. (2008). The domain of phonetic encoding in apraxia of speech: Which sub-lexical units count? *Aphasiology*, *22*(11), 1230–1247. https://doi.org/10.1080/02687030701820402

Zusammenfassung

Das Verstehen und die Produktion menschlicher Sprache sind komplexe Vorgänge, bei denen viele Faktoren und Prozesse zusammenspielen. Dennoch führen Sprachnutzer*innen sie zum größten Teil fehlerfrei, mühelos und ohne zeitliche Verzögerung aus. Um diese Effizienz zu erreichen, macht der menschliche Sprachverarbeitungsmechanismus sich sprachliche Prinzipien und Regelmäßigkeiten zunutze.

Welche Regelmäßigkeiten das im Einzelnen auf den unterschiedlichen Ebenen der Sprachverarbeitung sind, ist noch nicht vollständig erforscht. Die vorliegende Arbeit konzentriert sich auf die sublexikalische Ebene. Gegenstand der Untersuchungen sind wortinitiale deutsche Konsonantencluster.

Die Dissertation untersucht, inwieweit Sprachnutzer*innen bei der Produktion und Perzeption dieser Cluster von (nahezu) universellen phonologischen Prinzipien des Silbenaufbaus oder von der spezifischen graduellen Phonotaktik ihrer Muttersprache geleitet werden. Beide Einflussfaktoren wurden in der psycholinguistischen Literatur schon mehrfach als relevant für einzelne Aspekte der Sprachverarbeitung beschrieben. In dieser Arbeit werden ihre Rollen für die Sprachproduktion und -perzeption systematisch miteinander verglichen, wobei muttersprachliche und fremdsprachliche Perzeption separat untersucht werden.

Dazu werden universelle phonologische Prinzipien und sprachspezifische Phonotaktik wie folgt operationalisiert: Als Maß für universelle phonologische Wohlgeformtheit wird das Sonority Sequencing Principle (SSP) herangezogen, welches besagt, dass die Laute einer Silbe zum

Silbenkern hin steigende Sonorität aufweisen müssen. Sonorität ist ein phonologisches Konzept, das weitgehend mit der Intensität eines Phonems oder dem Öffnungsgrad des Mundes bei der Produktion korreliert. In der Sonoritätstheorie werden Phonemklassen auf einer Skala hierarchisch angeordnet, wobei Vokale über die höchste Sonorität verfügen und Plosive über die geringste.

Als Maß für sprachspezifische graduelle Wohlgeformtheit dienen Type- und Token-Frequenz, also die Gebrauchshäufigkeit, der einzelnen Cluster im Deutschen in silbeninitialer Position. Hochfrequente (HF) Cluster gelten also als sprachspezifisch wohlgeformter als niedrigfrequente (NF). Diese Einteilung stützt sich auf Erkenntnisse der gebrauchsbasierten Linguistik.

Es werden folgende Hypothesen aufgestellt:

1) HF-Cluster haben einen Verarbeitungsvorteil gegenüber NF-Clustern, der sich in der Perzeption und in der Produktion in geringeren Fehlerraten niederschlägt. Außerdem werden sie häufiger fälschlicherweise produziert und perzipiert, wenn sie nicht Targets (d.h. in der Produktion das Ziel der Sprachplanung, in der Perzeption das tatsächliche akustische Signal) sind.

2) Cluster, die das SSP befolgen, haben einen Verarbeitungsvorteil gegenüber Clustern, die gegen das SSP verstoßen, was sich ebenfalls in den jeweiligen Fehlerraten widerspiegelt.

3) Der Einfluss von Frequenz ist größer als der von Sonoritätsprinzipien, sodass im Falle entgegengesetzter Vorhersagen die frequenzbasierten zuverlässiger sind.

Nach einer kurzen Einführung in die Gesamtthematik (Kapitel 1) stellt Kapitel 2 die gebrauchsbasierte Linguistik vor, welche den theoretischen Hintergrund für die frequenzbezogenen Hypothesen bildet. In der gebrauchsbasierten Linguistik wird nicht zwischen Grammatik und Sprachgebrauch unterschieden, sondern angenommen, dass grammatikalisches Wissen auf der Kenntnis des Sprachgebrauchs basiert und von letzterem kontinuierlich beeinflusst wird. Sämtliches Regelwissen entstammt Generalisierungen anhand einer Vielzahl einzelner Gebrauchsereignisse und bedarf lediglich allgemeiner kognitiver Prozesse. Es wird daher davon ausgegangen, dass die Sprachnutzung einen Einfluss auf ihre Strukturen, mentale Repräsentationen und Verarbeitung hat. Insbesondere spielt die Frequenz von sprachlichen Strukturen eine wichtige Rolle für deren mentale Repräsentationen und spätere Verarbeitung. Je häufiger eine sprachliche Struktur benutzt wird, desto stärker verfestigt sie sich mental und wird infolgedessen leichter abrufbar. Darüber hinaus werden durch so genanntes Chunking und Entrenchment Sequenzen, die aus mehreren Elementen bestehen, als Einheit gestärkt und wachsen zusammen. Während dieses Phänomen bisher hauptsächlich in Bezug auf Mehrwortphrasen untersucht wurde, wird in der vorliegenden Dissertation erörtert, ob und inwieweit es sich auf die sublexikalische Ebene übertragen lässt. Die Annahme ist dabei, dass häufig aufeinanderfolgende Konsonanten ebenso Entrenchment unterliegen und Chunks bilden können, wie dies bei Mehrworteinheiten der Fall ist.

Kapitel 3 geht auf verschiedene Prinzipien für die Abfolge von Konsonanten ein und diskutiert Konsonantencluster als potentielle Einheiten in der Sprachverarbeitung. Zunächst wird die Bedeutung von sprachspezifischer Phonotaktik beleuchtet, wobei zwischen kategorischer und gradueller Phonotaktik unterschieden wird. Kategorische Phonotaktik umfasst absolute Regeln für die Zulässigkeit von tautosyllabischen Phonemfolgen, während gemäß gradueller Phonotaktik Phonemfolgen in Abhängigkeit von ihrer Frequenz mehr oder weniger wohlgeformt sein können. Für die graduelle Phonotaktik spielt also die Verteilung z.B. von Konsonantenclustern in einer bestimmten Sprache eine Rolle. Im Kontrast zu dieser sprachspezifischen Wohlgeformtheit steht eine universelle, auf Sonorität basierende Wohlgeformtheit von

Konsonantenclustern. Auf Basis des SSP (s.o.) können sie dichotom in wohlgeformt (d.h. dem SSP entsprechend) und nicht-wohlgeformt (d.h. gegen das SSP verstoßend) unterteilt werden. Das SSP besagt, dass initiale Konsonantencluster aus Plosiv und Liquid (z.B. /pl/) wohlgeformt sind, weil die Sonorität vom Plosiv zum Liquid und vom Liquid zum Vokal ansteigt, während Cluster aus Frikativ und Plosiv (z.B. /ʃt/) nicht wohlgeformt ist, da die Sonorität von /ʃ/ zu /t/ nicht ansteigt. Ob sie fällt oder gleich bleibt, hängt von der zugrunde gelegten Sonoritätsskala ab. Während Clements (1990) Plosive und Frikative als einheitliche Klasse der Obstruenten behandelt, stellen sie bei Selkirk (1984) zwei verschiedene Klassen mit unterschiedlicher Sonorität dar. In dieser Dissertation wird die feinere Unterteilung vorgenommen, nach der Frikative über eine größere Sonorität verfügen als Plosive. Das bietet die Möglichkeit, die Verarbeitung von Plosiv-Frikativ-Clustern und Frikativ-Plosiv-Clustern miteinander zu vergleichen, da beide im Deutschen wortinitial vorkommen. Darüber hinaus bietet das Sonority Dispersion Principle (SDP) eine graduellere Abstufung der Wohlgeformtheit. Es besagt, dass die Sonorität zum Silbenkern hin maximal ansteigen und danach minimal abfallen soll. Entsprechend sind unter den SSP-konformen silbeninitialen Konsonantenclustern diejenigen mit der größten Sonoritätsdistanz am besten und diejenigen mit der geringsten am schlechtesten. In Kapitel 3 werden darüber hinaus zwei Alternativen zu den Sonoritätsprinzipien vorgestellt, die ebenfalls einen Verarbeitungsvorteil für bestimmte Konsonantencluster erklären können, ohne auf deren Frequenz zurückzugreifen. Auf der einen Seite ist das die Net Auditory Distance (NAD), die im Gegensatz zur Sonorität Konsonanten numerische Werte nicht nur auf Basis der Artikulationsart, sondern auch des Artikulationsortes zuweist und stärker auf artikulatorischen und perzeptiven Beobachtungen basiert. Auf der anderen Seite sind es generalisierte Frequenzen, die auf natürlichen Klassen von Konsonanten beruhen. Das bedeutet, dass beispielsweise alle aus Plosiv und Liquid bestehenden Cluster eine Klasse bilden, deren Frequenz sich aus den addierten Frequenzen ihrer Mitglieder ergibt. Es wird auf Konsonantencluster als potentielle Einheiten in der Sprachverarbeitung eingegangen, die in der bisherigen Forschung wenig Aufmerksamkeit bekommen haben. Die 16 Cluster, die in allen Experimenten dieser Dissertation als Testcluster dienen, werden vorgestellt (/ts/,[1] /ʃt/, /ʃp/, /tr/, /kr/, /ʃl/, /fl/, /ʃm/, /pl/, /ʃn/, /sk/, /ps/, /sl/, /tʃ/, /ks/ und /sp/).

Kapitel 4 führt in das Thema der sublexikalischen Sprachperzeption als Grundlage für die beiden Perzeptionsexperimente ein. So unterscheiden sich verschiedene Konsonantenklassen in der Wahrnehmbarkeit ihrer akustischen Merkmale (acoustic cues), die darüber hinaus durch die Abfolge der Konsonanten bedingt ist. Plosive beispielsweise enthalten nur schwache (segment-)interne Cues, sodass ihre Identifizierung stark von externen Cues wie den Formantenübergängen in den folgenden Laut abhängt. Diese externen Cues werden am besten von Vokalen getragen, sehr schlecht von Phonemen mit schwach ausgeprägter Formantenstruktur. Frikative dagegen verfügen über sehr gute interne cues und sind daher kaum auf die angrenzenden Laute angewiesen. Exemplarisch werden zwei konnektionistische Modelle der Sprachwahrnehmung vorgestellt, das Neighbourhood Activation Model (NAM, Luce & Pisoni, 1998) bzw. seine Computerimplementierung PARSYN (Luce, Goldinger, Auer et al., 2000) und die Adaptive Resonance Theory (ART, z.B. Grossberg, 1976). NAM zufolge spielen für die Worterkennung hauptsächlich die Qualität des Stimulus, seine Unterscheidbarkeit von anderen Lexemen sowie seine Frequenz im Vergleich zu allen anderen aktivierten Kandidaten eine Rolle. In ART wird dagegen die Bedeutung von erlernten Erwartungen betont, die von Erfah-

[1] Obwohl es sich bei /t͡s/ aus sprachstruktureller Sicht im Deutschen um eine Affrikate handelt, wird es hier als Konsonantencluster behandelt, um als Vergleichsfall für die wesentlich niedrigerfrequenten Cluster /ps/ und /ks/ mit analoger Struktur zu dienen.

rungen geprägt sind. Worterkennung beruht demnach auf einem Abgleich dieser Erwartungen mit dem auditiven Input.

Auf dieser theoretischen Grundlage aufbauend, wird in Kapitel 5 die Perzeption von Konsonantenclustern durch erwachsene Muttersprachler*innen des Deutschen untersucht (L1-Perzeption). Die Hypothese ist, dass HF-Cluster häufiger korrekt perzipiert werden als NF-Cluster und dass sie eine höhere Rate an falschen Alarmen haben, also fälschlich anstelle der NF-Targets verstanden werden. Außerdem wird davon ausgegangen, dass SSP-Befolgung Clustern einen analogen Verarbeitungsvorteil verschafft, dass also SSP-konforme Cluster besser perzipiert werden und gegen das SSP verstoßende Cluster perzeptiv hin zu konformen Clustern korrigiert werden. Allerdings wird angenommen, dass sprachspezifische Erfahrung eine größere Rolle für die Perzeption spielt und Sonorität hauptsächlich auf die Perzeption von NF-Clustern einen Einfluss hat.

Versuchspersonen hörten mit Konsonantenclustern beginnende Pseudowörter, die in Stimmengewirr eingebettet waren, und transkribierten die Stimuli frei. Die Auswertung der Transkriptionen zeigt, dass HF-Cluster tatsächlich signifikant häufiger korrekt erkannt wurden als NF-Cluster. Außerdem wurden sie teilweise anstelle des eigentlichen (meist NF-) Targets gehört, bildeten also in Missperzeptionen öfter das Perzept als NF-Cluster. Der Effekt des SSP war indes entgegen der Annahme negativ: Gegen das SSP verstoßende Cluster wurden häufiger korrekt identifiziert als SSP-konforme Cluster. Dieses überraschende Ergebnis lässt sich auf den Einfluss von zwei Cluster-Gruppen zurückführen. Zum einen wurden die gegen das SSP verstoßenden Cluster aus Sibilant und Plosiv sehr zuverlässig erkannt, zum anderen hatten die SSP-konformen Plosiv-Sibilant-Cluster sehr hohe Fehlerraten. Letzteres ist teilweise dem perzeptiven Nachteil von initialen Plosiven geschuldet, wie die hohe Fehlerrate von /pl/ (und ferner /kr/ und /tr/) zeigt. Diese Beobachtungen zeigen, dass in der Perzeption von Konsonantenclustern neben rein akustischen Faktoren die Erfahrung mit sprachspezifischen phonotaktischen Verteilungen eine Rolle spielt, das universelle SSP hingegen nicht.

Es stellt sich allerdings die Frage, ob das auch für die fremdsprachliche (L2) Perzeption gilt oder ob diese stärker durch universelle Faktoren wie das SSP geprägt ist. Außerdem könnten dort die L1-Frequenzen zusätzlich zu den zielsprachlichen eine Rolle spielen. Um das herauszufinden, wurde das Experiment mit australischen fortgeschrittenen Lernenden des Deutschen wiederholt (Kapitel 6). Die Ergebnisse waren denen aus der L1-Perzeption sehr ähnlich, der Frequenzeffekt sogar etwas stärker ausgeprägt. Das zeigt, dass die Lernenden in der Lage sind, ihre Perzeption durch die Wahrscheinlichkeit der Konsonantencluster in der Zielsprache (ihrer L2) zu steuern, anstatt sich von der in ihrer L1 leiten zu lassen. Letztere zeigte keinen Haupteffekt, interagierte allerdings mit dem Effekt der L2-Frequenzen: Der Effekt der deutschen Frequenzen war am stärksten ausgeprägt für Cluster mit einer sehr niedrigen Frequenz (bis hin zu phonotaktischer Illegalität) im Englischen; Cluster mit hoher Frequenz im Englischen wurden unabhängig von ihrer Frequenz im Deutschen gleich gut erkannt. Die Lernenden sind also nicht vollständig unbeeinflusst von der Phonotaktik ihrer L1, sondern können die L2-Verteilungen besser erwerben, wenn sie unvoreingenommen durch jene in der L1 sind. Das Ausbleiben eines SSP-Effektes spricht dafür, dass selbst im Falle geringerer Erfahrung mit der Zielsprache die Perzeption nicht durch universelle sonoritätsbasierte Präferenzen des Silbenaufbaus gelenkt wird. Allerdings kann das nicht auf alle universellen Strukturprinzipien verallgemeinert werden, da die L2-Hörer*innen im Gegensatz zu den L1-Hörer*innen einen Einfluss von NAD zeigten: Cluster mit einer größeren NAD konnten sie zuverlässiger erkennen als solche mit einer geringeren. Das ist ein Indikator dafür, dass universelle phonologische Prinzipien in der L2-Perzeption einen höheren Stellenwert haben als in der L1-Perzeption – allerdings nur, wenn sie auf einem psychoakustisch realistischen Maß mit einem ausreichenden Feinheitsgrad basieren.

Kapitel 7 gibt, basierend auf bisheriger Forschung, einen Überblick über relevante Prozesse und Faktoren in der sublexikalischen Sprachproduktion und erörtert, inwiefern Versprecher Hinweise auf diese geben können. Lexikalische und sublexikalische Frequenz, strukturelle Komplexität, Ähnlichkeit zwischen Items und Wiederholung haben sich als einflussreich erwiesen. Außerdem wird die Spreading Activation Theory (Dell, 1986) vorgestellt, ein Modell der Sprachproduktion, dessen Hauptprinzip die Ausbreitung von Aktivierung in einem Netzwerk von Knoten auf unterschiedlichen Ebenen (Morphem, Silbe, Silbenkonstituente, Phonem, Merkmal) ist. Die Spreading Activation Theory ist in der Lage, korrekte Sprachproduktion sowie Versprecher zu modellieren. Phonologische Versprecher sind ihr zufolge darauf zurückzuführen, dass ein Nicht-Target-Phonem aufgrund einer großen Anzahl an Verbindungen zu anderen Knoten (welche Aktivierung senden) oder einer Verwechslung auf einer höheren Verarbeitungsebene (d.h. weil es in der geplanten Äußerung mehrfach vorkommt) stärker aktiviert war als das Targetphonem.

Analog zu den Perzeptionsexperimenten wird in Kapitel 8 spezifisch die Rolle von Clusterfrequenzen und deren Befolgung des SSP in der Produktion untersucht. In einem Zungenbrecher-Experiment wiederholten Versuchspersonen auditiv präsentierte Stimulus-Paare, deren Silben jeweils mit zwei ähnlichen Konsonantenclustern begannen (z.B. /sloːn fliːm/), vier Mal hintereinander. Als ähnlich wurden Cluster dann definiert, wenn sie sich entweder in nur einem phonologischen Merkmal in einem der Konsonanten unterscheiden (z.B. /sl/ und /fl/ im Artikulationsort des Frikativs) oder wenn sie dieselben Konsonanten in umgekehrter Reihenfolge enthielten (so genannte Metathese-Paare, z.B. /ʃt/ und /tʃ/). Durch die sehr schnelle mehrfache Wiederholung dieser Stimulus-Paare (144 Schläge pro Minute) wurden Versprecher induziert. Wie in den Perzeptionsexperimenten zeigte sich ein fazilitierender Einfluss von Clusterfrequenz: HF-Cluster wurden häufiger korrekt produziert als NF-Cluster und wurden häufig anstelle von NF-Targets produziert. Allerdings war der Frequenzeffekt in der Produktion weniger stark ausgeprägt als in der Perzeption. Als wesentlich größer erwies sich dagegen der Einfluss der Stimulus-Zusammensetzung: Trat ein Cluster in einem Metathese-Paar auf, so hatte es eine signifikant erhöhte Fehlerrate; das war auch bei HF-Clustern der Fall. Das bedeutet, dass es nicht so sehr eine inhärente Schwierigkeit von Clustern ist, die ihre Verarbeitung erschwert, sondern die Schwierigkeit sich vielmehr aus der Kombination der Phoneme in der Sequenz ergibt. Hier zeigt sich der Wettbewerb zwischen einzelnen Konsonanten, der ausgeprägter ist, wenn diese aufgrund mehrfachen Vorkommens stärker aktiviert sind. Bezüglich der Sonorität zeigte sich ein ähnliches Bild wie in der Perzeption. Allerdings war der fazilitierende Effekt eines SSP-Verstoßes nicht signifikant, sondern nur eine Tendenz. Wird sonoritätsbasierter Silbenaufbau graduell anhand der Sonoritätsdistanz bestimmt, so zeigt sich eine signifikant erhöhte Fehlerrate für Cluster mit einer Distanz von 1 auf der Sonoritätsskala[2], während alle anderen Gruppen mit ähnlicher Genauigkeit produziert wurden. Dieser Unterschied ist auch in der Produktion (wie vorher schon für die Perzeption festgestellt) auf eine erschwerte Verarbeitung von Plosiv-Sibilant-Clustern zurückzuführen. Ebenfalls parallel zur Perzeption erwiesen sich Sibilant-Plosiv-Cluster dagegen als besonders fehlerresistent. Sie waren auch häufig das Ergebnis eines Versprechers.

Zusammengenommen sprechen die Ergebnisse dieser Dissertation dafür, dass sublexikalische Sprachproduktion und -perzeption zu einem gewissen Grad denselben Mechanismen und Einflussfaktoren unterliegen.

Insbesondere sprachspezifische phonotaktische Verteilungen beeinflussen – in Übereinstimmung mit gebrauchsbasierten Vorhersagen – die Verarbeitung in beiden Modalitäten. Die Frequenz von Konsonantenclustern hat sowohl in der Produktion als auch in der (muttersprach-

[2] Das trifft auf Plosiv-Sibilant-Cluster sowie Sibilant-Nasal-Cluster zu.

lichen sowie fremdsprachlichen) Perzeption einen Einfluss auf ihre Verarbeitung. HF-Cluster sind weniger fehleranfällig als NF-Cluster. Der stärkere Effekt in der Perzeption kann dadurch erklärt werden, dass dort zusätzlich zu der Automatisierung der Verarbeitung von HF-Clustern ein (halb-) bewusster *Bias* hineinspielt: In Situationen erhöhter Unsicherheit – wie im Fall des degradierten Stimulus – lassen die Hörer*innen sich stärker von ihrem Wissen darüber leiten, welche Phonemfolgen in einer gegebenen Sprache wahrscheinlich sind.

In allen drei Experimenten durchgeführte Vergleiche bezüglich des Einflusses von Type- und Token-Frequenzen haben gezeigt, dass Type-Frequenzen für die sublexikalische Verarbeitung von Konsonantenclustern das relevantere Frequenzmaß sind. Im Produktionsexperiment zeigten ausschließlich Type-Frequenzen einen Effekt, in den Perzeptionsexperimenten war der Effekt von Token-Frequenzen geringer als der von Type-Frequenzen.

Sonoritätsprinzipien haben dagegen keinen Einfluss auf die Verarbeitung von legalen (inklusive marginalen) Konsonantenclustern gezeigt. Sollten sie in der Sprachverarbeitung prinzipiell relevant sein, so wird ihre Wirkung mit dem Erwerb einer spezifischen Sprache und das damit einhergehende Einüben auch (aus Sonoritätssicht) ungünstiger Phonemfolgen außer Kraft gesetzt. Insofern bestätigen die hier vorgestellten Studien die Position gebrauchsbasierter Linguistik, dass die Nutzung von Sprache ihre mentale Repräsentation und spätere Verarbeitung entscheidend prägt.

Zwei einzelsprachunabhängige strukturelle Prinzipien haben sich allerdings als einflussreich erwiesen. Zum einen zeigte NAD einen fazilitierenden Effekt in der L2-Perzeption; zum anderen stellten sich sowohl in der Produktion als auch in der Perzeption Sibilant-Plosiv-Cluster als besonders stark (hinsichtlich Fehlerraten und falscher Alarme) und Plosiv-Sibilant-Cluster als besonders schwach heraus.

In dieser Dissertation konnte gezeigt werden, dass Konsonantencluster – eine bisher vernachlässigte sublexikalische Einheit der Sprachverarbeitung – Frequenzeffekten unterliegen, die mit denen anderer sprachlicher Einheiten vergleichbar sind. Darüber hinaus wurde deutlich, dass auf sublexikalischer Ebene dieselben Mechanismen von Aktivierung und Wettbewerb wirksam sind wie auf lexikalischer Ebene.

This book presents three experimental studies of sublexical speech processing that aim to answer the question how universal principles of syllabic wellformedness like the Sonority Sequencing Principle, on the one hand, and experience with language-specific phonotactic distributions, on the other hand, influence perception and production of German initial consonant clusters.

Effects of consonant cluster frequency and sonority sequencing are compared in a) an experiment of pseudoword identification in noise with native German listeners, b) a parallel experiment with Australian learners of German, and c) an experiment of pseudoword repetition in a tongue twister paradigm. The results of these experiments have implications for the roles of universal vs. language-specific sound sequencing preferences in sublexical processing. The non-native listening experiment moreover allows for a comparison of the roles of native-language vs. target-language cluster frequencies in speech perception.

Against the background of usage-based linguistics and connectionist models of speech perception and production, this book expounds the mechanisms at work during the processing of consonant clusters. It discusses their role as potential units in speech perception and explores possibilities for consonant clusters having their own mental representations. A striking parallel is revealed between perception and production of consonant clusters in terms of which factors and principles facilitate or hamper their processing. It is one of the first works to demonstrate frequency effects for consonant clusters.

Sophia Wulfert studied General Linguistics at the Technische Universität Berlin and Scandinavian Studies at Humboldt-Universität zu Berlin. From 2015 to 2018, she was a member of the doctoral research training group "DFG GRK 1624: Frequency effects in language" at the University of Freiburg. This book is a revised version of her dissertation, which she defended in December 2021.

Die Publikationsreihe NIHIN – New Ideas in Human Interaction – entstand 2010 und ist ein Kooperationsprojekt zwischen der Hermann Paul School of Linguistics (HPSL) und der Universitätsbibliothek Freiburg (UB).

NIHIN bietet eine moderne, frei zugängliche Plattform für wissenschaftliche Essays erfahrener WissenschaftlerInnen sowie Prädikatsdissertationen, Textsammlungen zum Thema Sprache in der Interaktion und multimodale Sprachkorpora.

eucor
The European Campus

Universität
Basel

UNI
FREIBURG

ISBN 978-3-928969-97-0

9 783928 969970 >