Alona Kononenko-Szoszkiewicz

152522

Exploring (mor)phonotactic patterns: A comparative study of selected Germanic and Slavic languages

Badanie wzorców (mor)fonotaktycznych: studium porównawcze wybranych języków germańskich i słowiańskich

Rozprawa doktorska napisana

na Wydziale Anglistyki

Uniwersytetu im. Adama Mickiewicza w Poznaniu

pod kierunkiem prof. dr hab. Katarzyna Dziubalska - Kołaczyk

oraz emer. o. Univ.-Prof. Mag. Dr. Ulrich Wolfgang Dressler

Poznań, 2023

I dedicate this work to my mother.

Thank you for believing in me and letting me find my way 2000 km. away from home.

.

ACKNOWLEDGEMENTS

The order of the people I mentioned here is chronological, as I came to Poland in 2016 and became a student at Adam Mickiewicz University, Poznan. First and foremost, I would like to express my gratitude to my parents, Ela and Slava, for always believing in me and supporting me in all my endeavors. Next, I would like to thank Lukasz, my dear friend and life partner, for always having so much trust in me and encouraging me to pursue my academic career through my Master's and PhD studies. Without your tremendous support and faith, I would not be where I am today. I would also like to thank my Polish family, Renata and Krzysztof, who have always supported and encouraged my academic career.

I would like to express my deepest respect and gratitude to my supervisors and academic advisors, Prof. Katarzyna Dziubalska-Kołaczyk and Prof. Wolfgang Ulrich Dressler. "Szanowna Pani Professor," as I used to address you, you are my role model and teacher. Thank you for guiding me over the past seven years and for all the inspiring discussions and conversations we had during our seminars. Working with you was a great privilege.

Dear Ulli, I never imagined that I would have the opportunity to work with such an esteemed linguist. Thank you very much for your support, cooperation, and valuable insights during my studies. I greatly appreciate the time we spent together and our numerous conversations during our lunches in Vienna.

` Last but not least, to Misia, my greatest motivator and the joy of my life, who entered our world two years ago. You already speak three languages: Polish, Ukrainian, and Russian. One day, when you learn to read in English, I hope you will find time to see what your Mom has achieved.

OŚWIADCZENIE

Ja, niżej podpisany/a

Alona Kononenko-Szoszkiewicz

przedkładam rozprawę doktorską

pt. *Badanie wzorców (mor)fonotaktycznych: studium porównawcze wybranych języków germańskich i słowiańskich*

na Uniwersytecie im. Adama Mickiewicza w Poznaniu

i oświadczam,

że napisałam ją samodzielnie.

Oznacza to, że przy pisaniu pracy, poza niezbędnymi konsultacjami, nie korzystałem/am z pomocy innych osób, a w szczególności nie zlecałem/am opracowania rozprawy lub jej istotnych części innym osobom, ani nie odpisywałem/am tej rozprawy lub jej istotnych części od innych osób.

Jednocześnie przyjmuję do wiadomości, że gdyby powyższe oświadczenie okazało się nieprawdziwe, decyzja o wydaniu mi dyplomu zostanie cofnięta.

Poznań, 15.06.2023 Alona Kononenko-Szoszkiewicz

(miejscowość, data) (czytelny podpis)

Table of contents

List of research articles

The present PhD project comprises four thematically related research articles:

Research article 1 (Dressler et al., 2019)

Dressler, Wolfgang U., Alona Kononenko, Sabine Sommer-Lolei, Katharina Korecky-Kröll, Paulina Zydorowicz & Laura Kamandulytė-Merfeldienė. 2019. "Morphological richness, transparency and the evolution of morphonotactic patterns", *Folia Linguistica Historica, Folia Linguistica* 40(1), 85-106.

Research article 2 (Dressler & Kononenko-Szoszkiewicz 2019)

Dressler, Wolfgang U. and Alona Kononenko-Szoszkiewicz. 2019. "Main Differences Between German and Russian (Mor)phonotactics", In.: Wrembel, M., Kiełkiewicz-Janowiak, A., & Gąsiorowski, P. (eds.) *Approaches to the Study of Sound Structure and Speech: Interdisciplinary Work in Honour of Katarzyna Dziubalska-Kołaczyk (1st ed.)*. Routledge.

Research article 3 (Dressler & Kononenko-Szoszkiewicz 2021)

Dressler, Wolfgang U. and Alona Kononenko-Szoszkiewicz, 2021, "German phonotactic vs. morphonotactic obstruent clusters: a corpus linguistic analysis ", In.: W. U. Dressler, B. Calderone, S. Sommer-Lolei, K. Korecky-Kröll (eds.) *Experimental, Acquisitional and Corpus linguistic Approaches to the Study of Morphonotactics*, Austrian Academy of Sciences Press.

Research article 4 (Kononenko-Szoszkiewicz 2023)

"(Mor)phonotactics of Ukrainian. The study of word-initial consonant clusters", 2023, [accepted for publication in Italian Journal of Linguistics].

List of abbreviations

- **B&B –** the Beats & Binding model of phonology (Dziubalska-Kołaczyk, 2002)
- **NAD** – the Net Auditory Distance principle (Dziubalska-Kołaczyk, 2009)
- **MOA** – Manner of Articulation
- **POA** – Place of Articulation
- **S/O** – sonorant/obstruent
- **CV** – consonant-vowel sequence
- **CC** – consonant cluster
- **CA** – conversation analysis
- **DA** – discourse analysis
- **NLP** – natural language processing

List of figures

Part 1: Introduction

This section provides explanations of the terms that are essential to the current doctoral research project, namely phonotactics, and morphonotactics. Further, it is expanded by the overview of corpus phonology and bibliometric analysis of the literature relevant to the topic. The subsequent sub-sections provide a concise description and the primary objectives of the thematic articles that form part of this PhD thesis.

1.1. Defining key concepts

Languages are generally classified based on the sets of phonemes, but the number of sounds is fixed in every language. Although there are many possibilities of sounds combining with each other, the number of combinations is limited to the language. So, phonotactics studies permissible sound sequences in a particular language. The term 'phonotactics' was coined in 1954 by American linguist Robert P. Stockwell and originated from a combination of two Greek words meaning "sound" + "arrange" (cf. Hill 1958). Research on phonotactics involves analyzing the distribution of sounds in a language, identifying the sound combinations that are permitted or prohibited, and exploring how phonotactic patterns vary across languages. It examines the rules and constraints that govern the combination of phonemes. Phonotactic restrictions are widely acknowledged to be central to understanding the structure and patterning of language. While many of these constraints are influenced by the articulatory properties of speech sounds, some are primarily determined by the idiosyncrasies of a given language. Such limitations are often considered arbitrary, as they are not based on any inherent properties of the sounds themselves. When a consonant sound is followed by another consonant, they create a consonant cluster. For instance, in the English language, there is a well-known phonotactic constraint that forbids the occurrence of a nasal following a stop sound at the beginning of a word (Riitta 2005). A stop followed by a nasal, as in word-initial consonant cluster /gn/, is a permissible sound combination in Polish, Russian, and Ukrainian, like in the word 'гном' gnome, yet it violates phonotactics of English, and thus such sound arrangement is not possible.

Morphonotactics is a new field of research proposed by Dressler & Dziubalska-Kołaczyk (2006) to cover the interaction between phonotactics and morphotactics. According to the authors, the convergence between morphonotactics and phonotactics happens when morphological operations produce phonotactic sequences which already exist across morpheme boundaries. Although the morphonotactic patterns of language present a relatively new area of investigation, morphotactics refers to the first of Trubetzkoy's (1931) three parts or tasks of morphonology. According to Trubetzkoy, morphonology consists of the study of the phonological structure of morphemes, the study of combinatory sound changes undergone by morphemes in contact, and the study of sound alternation series serving a morphological function. Dressler (1985; 1996) defines morphonology as an area between morphology and phonology. As claimed by Trubetzkoy, only languages without morphology can do without morphonology. Dressler & Dziubalska-Kołaczyk (2006) also support this statement by saying that morphonology is primarily essential for the languages approaching the ideal inflecting-fusional type. Besides, Trubetzkoy stated that morphonology might play an essential role in providing a comprehensive description of languages concerning their linguistic typology. As specified by the author, the main task of morphonology is to identify the possible sound structures of different morpheme types (endings, prefixes, suffixes, etc.).

The division of morphonotactic and phonotactic clusters was proposed by Dressler & Dziubalska-Kołaczyk (2006). The former only appears in morphologically complex words, while the latter can only be found in morphologically simple forms and never result from morphological processes. An example of a morphonotactic cluster in English is /md/, which only occurs in complex words like "seemed" and "doomed" but never in simple forms. On the other hand, the /mp/ cluster in English words such as "camp" and "chimp" is purely lexical and only found in morphologically simple words since it does not result from any morphological operations. Sometimes the same clusters can be ambiguous, either phonotactic or morphonotactic, depending on the environment. For instance, the word-final consonant cluster /-nd/ in a word *open+ed* is morphonotactic due to the morpheme boundary of the word coda and suffix -ed. Still, in the word *hand,* it is phonotactic or lexical, as no morphological operations are involved. In order to avoid ambiguity, the authors proposed a scale of deviation from purely morphonotactic to phonotactic clusters:

- exclusively morphologically motivated clusters (e.g., the word- final clusters /-fs, -vz/ as in laughs, wife's, loves, wives);
- morphologically motivated as a strong default (e.g., the word-final clusters /-ts, -dz/, as in cats, kids, which occur in just a few mono- morphemic words, i.e., waltz, grits, adz(e), and the loan-words quartz, kibbutz);
- morphologically motivated as a weak default (e.g., the word-final cluster /-ks/, as in docks, lacks, which occurs in mono-morphemic Latinate words such as tax, sex, box, flux, fix, six);
- morphologically motivated in the majority (e.g., in Italian, the word-initial clusters /zl-, zm-, zn-, skw-/ contain a morpheme boundary in the majority of cases, as in s+leale 'disloyal' vs. slang, s+membrare 'dismember' vs. smog, s+naturare 'denaturate' vs. snack, s+qualificare 'disqualify' vs. squadra 'team');
- morphologically motivated in the minority (e.g., in Polish, the word-initial clusters /sp-/, /sk-/, /st-/ and /vj-/ contain a morpheme boundary in the minority of words they head, as in s+palić 'burn' vs. spać 'sleep', s+kończyć 'finish' vs. skoczyć 'jump', s+toczyć 'tumble' vs. stać 'stand', w+jechać 'drive in' vs. wiem 'I know').

Slavic languages are supposed to have a significant number of morphonotactic and phonotactic consonant clusters due to their rich morphology. Still, their asymmetric distribution within the word differs from Germanic languages. This difference is reinforced by morphology because Ukrainian and Russian have mono-consonantal prefixes, and German and English have mono-consonantal suffixes. This is why Germanic languages have developed a much richer morphology in medial and final positions, contrary to Ukrainian and Russian, which have many more consonant clusters in word-initial position. In English and German, in the medial position, there are many morphonotactic and phonotactic consonant clusters due to morphological processes such as derivation or compounding. The main word-internal difference is that in English and German, compounding increases the number of consonant clusters, whereas in Russian and Ukrainian decreases them due to vowel insertion.

1.1.1. Beats-&-Binding phonotactics

There are many alternative sonority-based models which can be used for the evaluation of syllable structure. For the purposes of the present research, an alternative approach for cluster evaluation based on the universal model of phonotactics constructed within the Beats-&-Binding phonology model was applied. Such a choice has been motivated by the fact that this model goes beyond purely sonority-based models and is not attached to any of the traditional syllabification models. By taking into account the perceptual contrast between beats and non-beats, it allows to evaluate cluster preferability and establish a hierarchy of the preferences of clusters from the most preferred (unmarked) to the least (marked). Perceptual contrast of the consonants is measured by means of the Net Auditory Distance principle (Dziubalska-Kołaczyk 2009).

According to the model, every beat (typically vowel) and non-beat (always consonants including glides) are connected to each other by the bindings, which are organized due to the phonotactic preferences such as the preference for a trochee, for the vocalic beat, or the alternation between beats and non-beats. Moreover, articulatory preferences are taken into account for the cluster evaluation which govern the choice of segments with the perceptual sonority-cued preferences and thus codetermine the shape of phonotactics. Therefore, beats and non-beats have direct phonetic correlates both in production and perception.

Phonotactic preferences are closely related to the notion of markedness. The concept of markedness goes back to the scholars of the Prague school. Trubetzkoy (1939) wrote that if two phonemes share the same set of features, except for one feature found in only one of the phonemes, this feature is the ‘mark’ and involves an extra articulatory gesture. Moreover, according to the author, markedness is closely related to articulatory complexity, the combinatory possibilities of sounds, phonological statistics, functional load, and neutralization. As a rule, onset clusters and coda clusters tend to be more marked than single consonant onsets and codas. The constraints that forbid these clusters can be considered as complex onset and complex coda (Prince and Smolensky 1993). Syllables that do not have codas are less marked than syllables that have a coda. This is due to the violation of the constraint prohibiting complex onsets and codas.

In modern meaning, the notion of markedness often refers to the measure of the naturalness of linguistic elements. However, the definition of markedness is greatly dependent on linguistic theories. For instance, from the point of view of Generative Phonology, Chomsky & Halle (1968) suggest that with the use of markedness, they have the machinery for making distinctions between more and less plausible rules in purely formal terms.

In the Beats-&-Binding phonology, markedness can be understood within the theory of universals, which are defined as the properties of the language that can be scaled from the most natural to the least, even though the preferences which are accountable for in one language may differ in another. Thus, the most preferred properties, in the present study phonotactic preferences, are understood as unmarked and natural, whereas those dispreferred can be considered marked and unnatural. As shown earlier, the alternating sequence of consonants and vowels leading to CV-syllable structure is perceived as universally preferred among many languages, and therefore it is unmarked. As pointed out by Dziubalska-Kołaczyk (2002), the perceptual contrast between segments should be sustained for clusters to survive, and this contrast was defined as the Net Auditory Distance Principle (hereafter The NAD Principle).

The NAD Principle was preceded by the Optimal Sonority Distance Principle (Dziubalska-Kołaczyk 2002), according to which “the sonority distances between the sounds should be optimally balanced.” A new model is based on manner of articulation (hereafter MOA), place of articulation (hereafter POA) as well as sonorant-obstruent distinction (S/O). The NAD principle defines cluster preferability in relation to the position in the word (initial, medial, and final). A cluster preferability is measured by the Net Auditory Distance calculator introduced by Dziubalska-Kołaczyk et.al. (2007, 2014). The definition of preferability is formulated as follows:

A cluster is preferred if it satisfies a pattern of distances specified by the universal phonotactic preference relevant for its position in the word.

Thus, the calculation for the word-initial cluster C1C2V is provided below:

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

$$\text{NAD CC} = |(\text{MOA1} + \text{MOA2})| + |(\text{POA1} + \text{POA2})| + 0/1$$

For instance, the NAD prediction for word-initial double consonant clusters prV in Polish according to the NAD calculator is as follows:

prV= |(5 - 2)| +|(1 - 2.3)| = |3| +|1.3| +1 = 5.3, so NAD CC=5.3

rV: |(MOA1 - MOA2)| =|2| - 0= 2, so NAD CV = 2

Thus, the preference NAD (C1, C2) > NAD (C2,V) is observed since 5.3 >2.0.

For word-initial triple clusters C1C2C3V, the prediction is the following:

NAD (C1,C2) > NAD (C2,C3) > NAD (C3,V)

The condition reads:

"For word-initial triple clusters, the NAD between the second consonant and the third consonant should be greater than or equal to the NAD between this third consonant and the vowel, and greater than the NAD between the second and the first consonant".

The NAD product was introduced to the calculator in order to receive a preferability index which is "a number denoting a degree to which a given preference is observed" (Dziubalska-Kołaczyk 2018). The formula for word-initial consonant clusters (C1C2V) is as follows:

NAD product = NAD C1C2 – NAD C2V

For instance, the NAD product for the word-initial double consonant cluster (C1C2V#) in English:

(1) pɹV

NAD pɹ =(5-2)+(2.6-1)+1=5.6

NAD ɹV=2+0=2

NAD product =5.6-2=3.6 (preferred cluster)

NAD product = NAD C1C2 – NAD VC1

The NAD product for the word-final double consonant cluster (VC1C2#) is the following:

(2) Vkt

NAD kt =0+(3.5-2.3)=1.2

NAD Vk =5+1=6

NAD product =1.2-6= -4.8 (dispreferred cluster)

The general prediction of the Beats-&-Binding model is that morphonotactic clusters are expected to be marked, since in this way they signal morphological boundaries. As claimed by Dziubalska-Kołaczyk (2009), the predictions of cluster preferences may account for language-specific phonotactics, its acquisition and change. Moreover, such an analysis of cluster preferability may explain the order of difficulty in the acquisition of phonotactic clusters, as it has been shown in the study on English and Polish phonotactics by Zydorowicz (2009).

1.1.2. Corpus phonology and bibliometric analysis

The cycle of publications that shapes this PhD dissertation includes the articles which use cross-linguistic methods of research. A multidisciplinary approach to studying phonotactics and morphonotactics is based on the application of corpus linguistic methods, i.e., corpus-based research. Corpus phonology is a modern multidisciplinary area of study that combines methods and theoretical approaches from phonology, diachronic and synchronic linguistics, phonetics, corpus linguistics, speech technology, information technology, computer science, mathematics, and statistics. It developed out of the need for modern phonological research to be embedded within a larger social, cognitive, and biological science framework. It is commonly acknowledged that a corpus-based methodology has the greatest potential when used to generate new linguistic hypotheses and uncover previously unknown linguistic phenomena due to the heuristic strength of corpus searches (Biber et al. 1994).

Historically, corpus linguistic methods were employed in phonology to carry out qualitative research on the distribution of a particular sound or on the ways by which it is realized, to discover language-specific variation and phonetic patterns. Corpus linguistic methods prove to be an invaluable tool in the study of sound systems since they involve using large language corpora, or collections of written, spoken, or transcribed language, as a source of data for phonological analysis. In addition to providing linguists with a wealth of data to analyse qualitatively, corpus linguistic methods also allow for the quantification of data. This is particularly useful in the study of phonology, as it allows for the comparison of sound patterns across languages, dialects, and contexts. As highlighted by Delais-Roussarie et al. (2014), these days, corpora are frequently utilized to propose statistical modelling and develop probabilistic grammars that account for phonological facts and variability (Pierrehumbert 2003a; 2003b) as well as to validate theoretical presumptions and to even replace the linguist’s intuition (Cori and David 2008; Durand 2009). Frisch (2012) makes a strong point about the use of speech databases and lexical corpora for the analysis of frequency, probability, and similarity in phonology corpus data. As noted by Durand et.al (2013), the utilization of corpus linguistics methods applied to phonological research is still in its infancy.

The validity of corpus-based linguistic theories heavily depends on the corpus's quality. To draw accurate and generalizable theoretical findings, it is generally agreed that a sample of primary material (written texts, audio or video recordings, or both) must be gathered that is maximally representative of the language or linguistic variety (Voormann et al. 2008). For instance, speech corpora are often collected from specific populations or geographical areas, making it difficult to generalize the findings to a broader population. Additionally, the amount of data available for a particular speech corpus may be limited, making it difficult to draw more reliable and comprehensive conclusions. Finally, the reliability of speech corpus can be influenced by the quality of the recordings, as well as the transcription process. Poor-quality recordings can contain background noise and other distortions. Thus, if the transcription process is not accurate or consistent, then the reliability of the data is compromised, which might be crucial specifically for the study of phonetics and phonology of a given language. Yet, as suggested in *The Cambridge handbook of English corpus linguistics*, most corpora are solely made up of written texts, and those which are a mixture of written and spoken data texts are still overwhelmingly written (Biber & Reppen 2015).

Phonetic and phonological aspects of the language are still quite rarely illustrated in corpus research. It might be related to the fact that there is a moderate number of available sources; more specifically spoken corpora are often not accessible for public use due to various reasons. The compilation of such corpora requires consent from the participants from further reuse of the data, it is greatly time-consuming and technically might be challenging to perform, and some researchers do not want to share their data. Yet the research on written types of corpora seems to be more accessible and easier to perform due to the great number of statistical tools which might facilitate the research. Therefore, overall research on spoken and written corpora might be disproportionately shaped.

To explore what is the actual scope of research that applies cross-linguistic methods, i.e., corpus linguistic methods used for the investigation of phonetic and phonological phenomena, I have conducted a bibliometric analysis of all publications from the Scopus database from the years 2010-2020. Scopus is the most complete repository of bibliographic citations. Bibliometric analysis is an increasingly popular tool for analyzing trends in linguistics and other scientific fields. It is a quantitative method of measuring the impact of research that uses bibliographic data, such as citations, key authors, topics, and trends in a particular field. The application of bibliometric analysis allowed me to gain insight into the development of a field over time, identify key contributors and the relative impact of their works, and determine emerging trends and areas of linguistic research that may be ripe for further exploration. Scoping review methodology has been applied to bibliometric analyses since it allows for identifying knowledge gaps, defining the scope of a body of literature, and investigating research conduct (Munn et al. 2018).

As Hilary Arksey & Lisa O'Malley (2005) noted, the scoping review typically includes the following five stages, which were applied to the present complementary research.

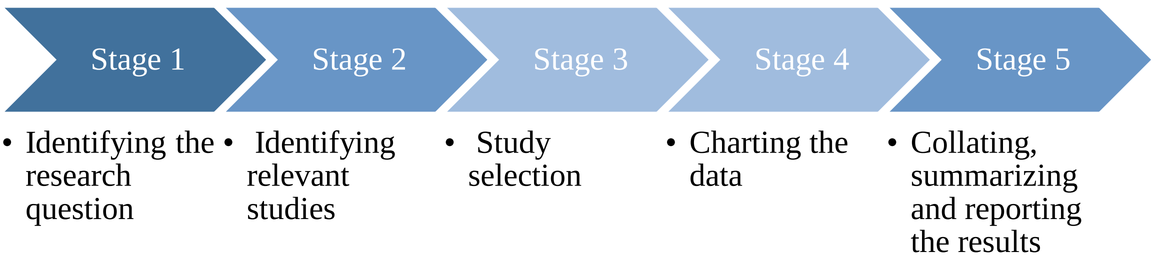


Fig. 1. Stages of the scoping review analysis

To explore what are the tendencies in corpus linguistic research and what is the role of phonology studies within this domain, the following research questions have been postulated:

1. Which types of corpora are predominately used in corpus linguistics?
2. Which languages are dominant in corpus linguistic studies?
3. What are the most common domains of linguistics that utilize corpus linguistics methods of research?

Scopus database allowed to compile the dataset of the publications relevant to the present research from 2010-2020. The dataset was structured as an Excel document. It included all relevant information for this study, precisely the document type, year of publication, authors, title, languages of publication, abstract, keywords, etc. A multifaceted search query was executed to retrieve records from the Scopus database (Appendix 1).

The selection of relevant studies was based on the co-occurrence of selected keywords, including the following: 'written corpus/corpora, 'spoken corpus/corpora'. The decision to choose broad keywords was made with the aim of retrieving a comprehensive set of research articles, despite the possibility of including some irrelevant ones. However, the application of scoping review methodology allowed the elimination of all irrelevant articles from the dataset.

The research was limited only to articles and book chapters on the relevant topic. The publications were exported to Excel and annotated manually according to the inclusion criteria, i.e., articles not relevant to the field of linguistics were excluded. The data cleansing also involved the removal of duplicates. The final dataset did not include books and conference papers/conference proposals.

The third stage of research was divided into two phases: screening I (screening by the title) and screening II (screening by the abstract).

The first phase of data preparation - screening by the title involved scanning the titles of the articles retrieved from the Scopus database and removing irrelevant articles as they did not meet the inclusion criteria. For instance, the articles related to medicine or law where the word ‘corpus’ has been used in its literal meaning originating from Latin denoting the human body. Another instance of the occurrence of the word ‘corpus’ has often been found in the texts related to religion when it was used in the phrase Corpus Christi. During this phase, the data has been controlled by checking authors and index keywords. Thus, the overall number of the records screened by the title is 2230 articles. After excluding all irrelevant publications, 1552 records were selected for analysis in the next phase (Figure 2).

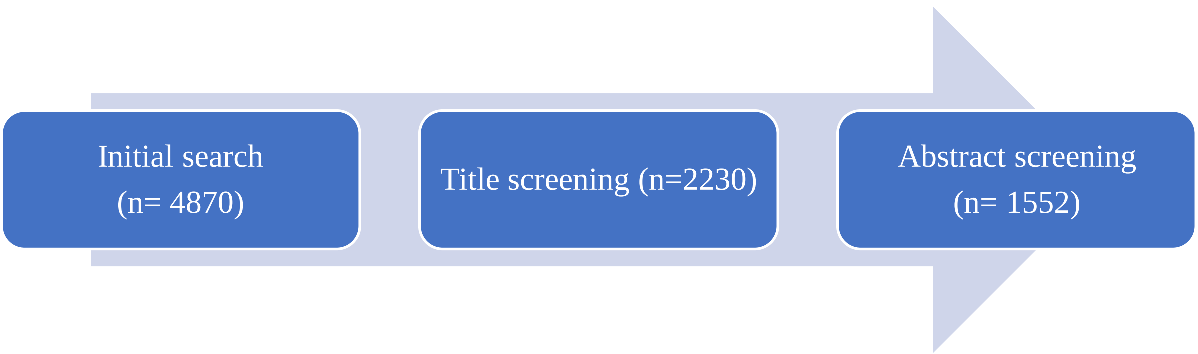


Fig. 2. Publications identified within the database

The second phase – screening by the abstract involved reading the abstracts of the articles selected in the first phase of the analysis and annotating them according to the eligibility criteria. The same inclusion guidelines have been followed. Additionally, there has been annotated the type of corpora used in the research (e.g., written, spoken, other) and the language studied in the research paper.

Upon completing the final dataset compilation, the acquired data allowed me to respond to the first and second research questions. In order to determine the domains of linguistic research that employ corpus linguistic methodologies for analysis and address the third research question, the VOSviewer software was employed to visualize bibliometric data and identify co-occurrence networks (van Eck & Waltman 2010).

In response to the first research question, findings indicate that in corpus linguistic studies, written types of corpora are the primary focus. Out of a total of 1552 publications that utilized language corpora for research purposes, 860 were based on written corpora, 575 on spoken corpora, 96 studies utilized both spoken and written corpora, and only 21 studies used corpora of sign language.

The second research question referred to the languages that are primarily the subject of investigation in corpus linguistics. As expected, the English language emerges as the predominant focus, given that the largest existing language corpora are The Corpus of Contemporary American English comprising over 560 million words (Davies 2008-), which is 5 to 6 times larger than the British National Corpus (comprising 100 million words). The data revealed that the overall number of languages that are a subject of corpus linguistic research is 212 As depicted in Figure 3, the size of each circle denotes the frequency of occurrence, while the interconnections between languages signify those that are most commonly studied in conjunction with each other. It is worth noting that, in VOSviewer the languages are presented in lowercase letters as a standard convention. The examples of language, along with their precise frequencies in the dataset, are provided in Appendix 2.

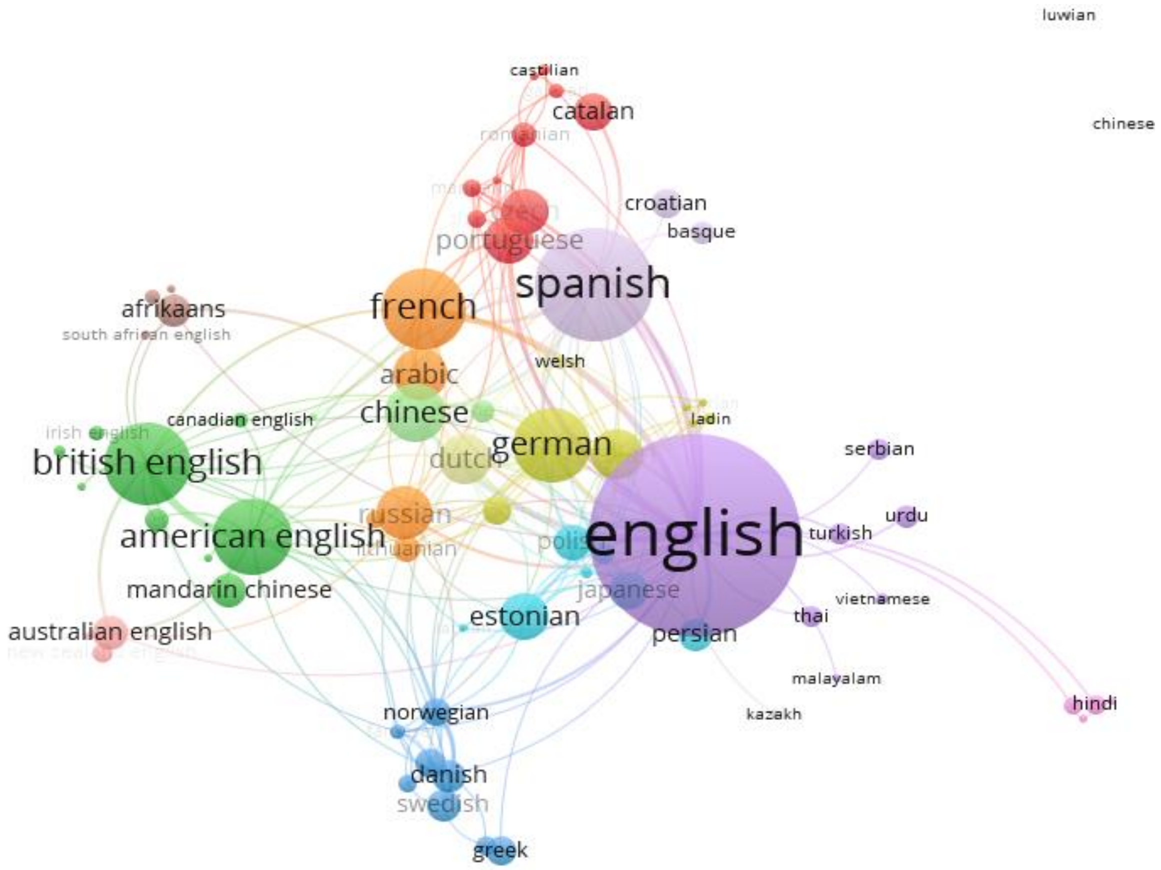


Fig. 3. Languages of the research in Scopus 2010-2020

Screening by the keywords allowed me to identify various research streams that characterize and define corpus linguistic tendencies as well as highlight the connections and relationships between them. The analysis revealed that there are four major domains of linguistics that utilize corpus linguistic research methods. As demonstrated in Figure 4, the strongest cluster is related to academic discourse (red), followed by speech recognition and natural language processing (green), language development (blue), and speech analysis (yellow).

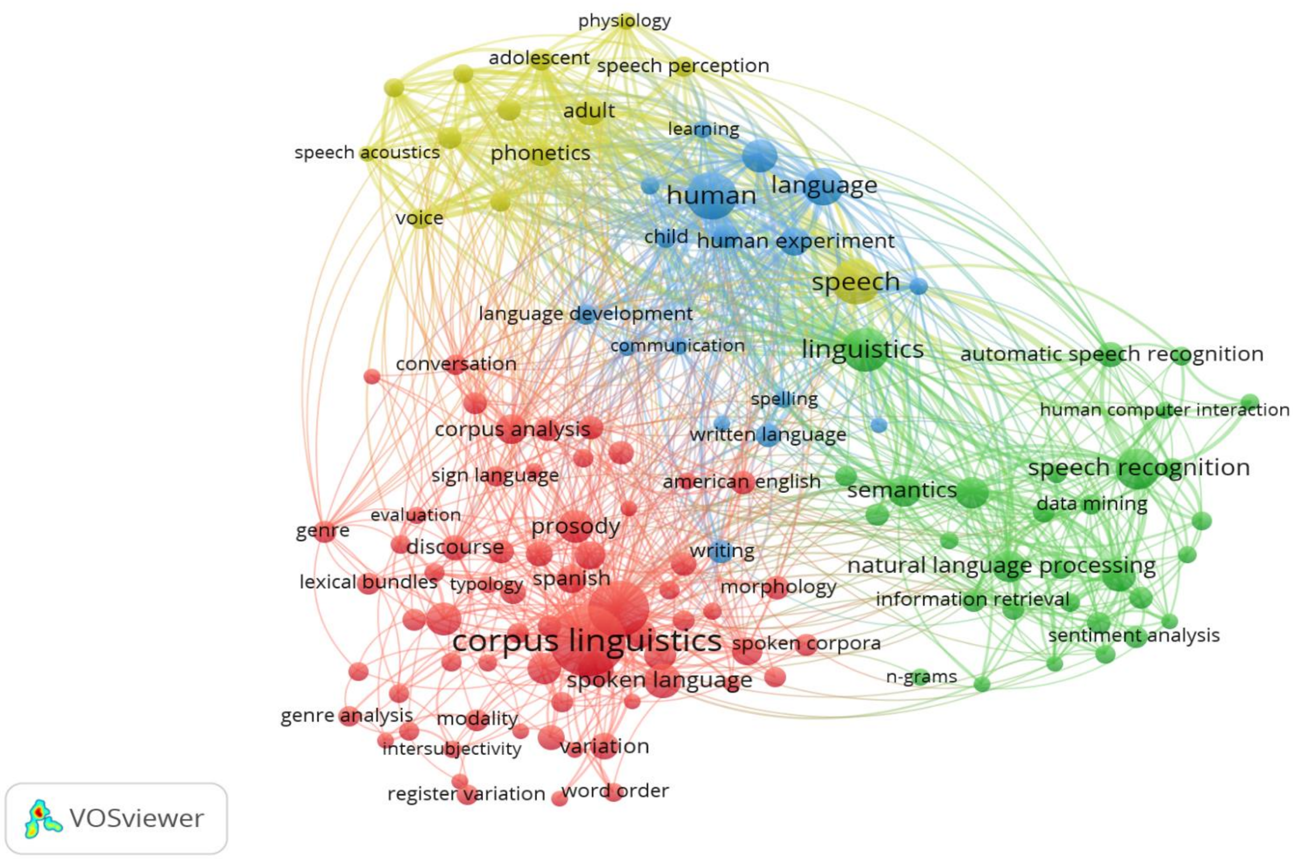


Fig. 4. Thematic landscape of corpus linguistic research

The largest area of linguistic research employing corpus methodologies is discourse studies. Corpus methodologies in discourse studies encompass a wide range of research traditions, including conversation analysis (CA) and discourse analysis (DA). While CA focuses on the organization and sequential structure of talk-in-interaction through the detailed examination of audio and video recordings, DA investigates the social, cognitive, and ideological aspects of discourse using qualitative and quantitative methods to analyze large corpora of spoken or written texts. Despite their different methodologies, CA and DA share some common ground, such as the importance of context, the co-construction of meaning, and the investigation of power dynamics, identities, and ideologies. Thus, the incorporation of corpus linguistic methodologies in discourse studies leads to more robust and nuanced understandings of language use in social contexts.

In both speech recognition research and natural language processing (NLP) domains, corpus linguistic methods play a crucial role in analyzing and understanding the intricacies of language. By facilitating the examination of phonetic, prosodic, and acoustic features within spoken language corpora, these methods enable researchers to model the variations and complexities inherent to spoken language, including accents, dialects, and speech disfluencies. This, in turn, contributes to the enhancement of speech recognition systems, rendering them more accurate and robust in handling diverse speech patterns and styles. Simultaneously, corpus linguistic techniques are employed in NLP to study linguistic phenomena at various levels, such as syntax, semantics, and pragmatics. Through the analysis of large corpora, NLP researchers can uncover the underlying structures and relationships between linguistic elements. These insights are then incorporated into the development of algorithms for tasks such as machine translation, sentiment analysis, and information extraction. Consequently, the application of corpus linguistic methods in both speech recognition and NLP research significantly advances our understanding of language use and its inherent complexities, ultimately contributing to the ongoing evolution of these interdisciplinary fields.

The third hub shows the application of corpus linguistic methodologies as an essential approach in the study of human language development, offering valuable insights into the multifaceted processes of language acquisition. For instance, in the context of first language acquisition, corpus linguistic methods facilitate the analysis of longitudinal child language data, elucidating the emergence and progression of linguistic features, including phonetic, morphological, syntactic, and semantic structures. These analyses contribute to a deeper understanding of developmental milestones, stages, and the impact of social, cognitive, and environmental factors on language acquisition.

The fourth hub focuses on the common theme related to the analysis of speech processing and speech perception. Both fields rely on the use of data, and together they offer a robust set of tools for understanding how we process and understand spoken language. Phonetics, speech acoustics, speech analysis, speech production measurement, and voice analysis are some of the techniques employed in studying the physiological and cognitive aspects of speech production and perception. These methodologies have been instrumental in investigating the physiological and acoustic features of speech production and the cognitive processes underlying speech perception and comprehension. Furthermore, corpus linguistic methods enable the study of the impact of individual differences, such as age and linguistic background, on speech perception and processing. Moreover, this trend is especially strong in sociolinguistics (e.g., Kendall 2013) and psycholinguistics (e.g., Meyer et al. 2016).

By and large, Scopus is an extensive database of scholarly research articles and scientific publications. It is often used by the scholars to search for and track academic literature, as well as to measure the impact of research. However, like any large database, Scopus has certain limitations one shall consider when using it. While it contains a large number of articles and publications, it is not a complete database of all existing scholarly literature on the specific topic. Therefore, some journals, particularly those in smaller or specialized fields, may not be included in the database. Thus, it is not fully comprehensive. For instance, with regards to morphonotactics, only seventeen papers are included in the Scopus database. As noted by Siversten (2014), there are disciplinary differences since many influential publishers of scholarly books, particularly in the social sciences and humanities, are not covered by Scopus compared to the publications in the fields of science, technology, and medicine (Hicks 2004; Nederhof 2006; Martin et al. 2010).

Additionally, Scopus has an emphasis on English-language publications, which may lead to an underrepresentation of non-English literature, as have been demonstrated in Figure 3. Another limitation of Scopus is that it may have some bias towards certain publishers or journals. This can skew the results of searches and metrics, making it appear that some research or researchers are more impactful than others. The deficiencies are mainly due to incomplete coverage of international journals, limited or no coverage of national disciplinary journals, and very limited coverage of peer-reviewed scholarly publications.

To sum up, the bibliometric analysis allowed us to gain insight into the research of corpus linguistic studies and define dominant trends in this field from the last decade. As the data shows, while the investigation of phonetics and phonology may not be overtly articulated in the titles of scholarly articles, the exploration of speech and sounds is fundamentally ingrained within the domain of corpus linguistics. Speech perception, speech recognition, and speech processing involve the analysis of the language from interdisciplinary perspectives. On the one hand, the traditional description of the sound systems of specific languages, and on the other, the application of tools and methods not traditionally utilized in language studies. Consequently, the availability of language corpora has enabled me to examine phonotactic patterns of the languages not merely descriptively but also quantitatively, which will be further elaborated on in the subsequent sections of this doctoral thesis.

1.2. Research article 1. Morphological richness, transparency and the evolution of morphonotactic patterns.

Identifying morphonotactics as a proper subfield of morphonology opens new perspectives for diachronic studies. Previous research on the evolution of morphonotactic clusters by Dressler et al. (2010) suggests that the historical origin of consonant clusters in terms of their formation over time depends on two factors. First, the language's morphology must have a certain level of complexity. Second, we anticipate that these types of clusters will only emerge due to the presence of complex phonotactics in the language. According to the authors, phonological vowel deletion is a common diachronic source for the development of morphonotactic consonant clusters. The authors provided evidence supporting their predictions by presenting a historical overview of the evolution of (mor)phonotactics, beginning with the reconstructed Proto-Indo-European language and modern Balto-Slavic languages. Bauman & Kaźmierski (2016) presented a mathematical model tested against Polish and English synchronic and diachronic language data. The study revealed that the evolutionary dynamics of cluster inventory depend on how the signaling function of morphonotactic clusters is compromised by the presence of lexical items that contain their internal morpheme counterparts.

Research article 1 (Dressler et.al. 2019) aimed to analyze data from multiple languages and examine the processes responsible for creating and changing consonant clusters. We expand upon prior research to the studies on language acquisition and present supporting evidence from language processing to formulate explanations. By surveying and categorizing these processes, we aim to understand better the formation and evolution of consonant clusters across languages.

This study focuses on the historical emergence and development of morphonotactic consonant clusters in various languages, including Germanic, Slavic, Baltic, Romance, and others. We investigate the impact of several morphological preference parameters, such as morphotactic and morphosemantic transparency/opacity and morphological richness of the selected languages. Furthermore, the paper identifies various diachronic processes that contribute to cluster formation, production, and change, including but not limited to vowel loss, Indo-European ablaut (and similar Arabic processes), affixation, compounding, metathesis, and final and consonant epenthesis. With regard to preference, six hypotheses have been tested, which explored the impact of preference parameters on the development of morphonotactic clusters.

Consistent with previous research, vowel loss was identified as the primary source for the formation of phonotactic consonant clusters in Germanic languages such as German and English, as well as in Slavic languages, more specifically, Russian, Ukrainian, and Slovak. However, Lithuanian morphonotactic clusters that arise due to vowel deletion occur only in imperatives in the word-final position. In Latvian many consonant clusters emerged in the unstressed final syllables.

Affixation is another mechanism for the creation of morphonotactic consonant clusters in word-initial and word-final positions, for instance, s- prefixation in Italian. A significant number of morphonotactic clusters which arise due to the prefixation can be found in Polish, Russian, Ukrainian, and Slovak. Other sources of cluster emergence presented in the article involved compounding, mainly in German and Lithuanian, metathesis in Polish, and epenthesis in word-final and word-medial positions in German and Polish.

The general tendency of languages is to avoid consonant clusters, and it has been postulated that the dispreference towards clusters is stronger in morphonotactic than in phonotactic clusters. It has been suggested that the morphologically richer Slavic languages have more morphonotactic clusters than the morphologically poorer Germanic and Romance language. The relationship between cluster preference and the type of consonant cluster has been a central point of research in other studies by the author, which are presented further in this cycle of publications. Based on cross-linguistic analyses and observations of language acquisition and processing, the authors agree that any disparities between the developmental trajectories of morphonotactic clusters and phonotactic clusters are subtle and difficult to identify.

To sum up, a current study presented an analysis of empirical evidence drawn from a diverse set of languages, mainly Germanic, Roman, and Slavic languages. Our investigation involved a thorough examination of the various phonological processes that contributed to the formation of morphonotactic consonant clusters. This approach allowed gaining a deeper understanding of the factors that influence the emergence and evolution of clusters in human language and opened new possibilities in the research on the cognitive and neural mechanisms that support their perception, production, and acquisition.

1.3. Research article 2. German phonotactic vs. morphonotactic obstruent clusters: a corpus linguistic analysis.

This article presents an analysis of the phonotactic structures of German presented in The Austrian Media Corpus, paying attention to morphological boundaries (Resch & Dressler 2017, Ransmayr 2018). For the first time, we demonstrated a typological characteristic of the German language regarding phonotactics and morphonotactics compared to previous studies on German, Polish, and English. Similarly to other Germanic languages, German is characterized by a relatively large inventory of consonants compared to its vowel sounds. In addition, it makes use of a wide variety of complex consonant clusters. To provide quantitative evidence of the type and token inventory, a corpus linguistic analysis was carried out.

Previous research on German phonotactics presents only a tentative analysis of the selected consonant clusters (Hyman 2007; Blevins 2007; Calderone et al. 2014, Hyman & Plank 2018). Current research characterizes German word-peripheral patterns of consonantal morphonotactics vs. phonotactics from phonological, morphological, typological, and corpus-linguistic perspectives. In our study, we follow a well-established framework of consonant cluster analysis, namely the Beats-and-Binding model of phonotactics (Dziubalska-Kołaczyk, 2002).

The purpose of corpus linguistic analysis is twofold. In usage-based linguistics, the frequency of the overall occurrence of a word or phrase and the difference between high and low frequencies are significant for determining how productive and valuable they are in language patterns. Therefore, we wanted to quantify the type and token ratio of all word-initial and word-final consonant clusters to check the productivity of German consonant clusters. Second, the generally accepted notion is that consonantal languages have more dispreferred consonant clusters than vocalic languages. To put this assumption into practice and evaluate the degree of preferences of consonant clusters in German we have applied the Net Auditory Distance principle (Dziubalska-Kolaczyk et.al. 2007, 2014). All consonant clusters were divided according to its status, phonotactic, morphonotactic or both.

The analysis revealed that in German the complexity, frequency, and typological diversity of word-final clusters are considerably greater than those of word-initial clusters. These findings are supported by empirical evidence and contribute to a more precise understanding of the phonological characteristics of Germanic languages. In the word-final position, there can be up to four consonants clustered together only of morphonotactic nature which predominantly occur in 2nd SG., 3rd SG. or past participle. According to the corpus data for word-final quadruple clusters the number of occurrences is in direct relation to the type frequency although with a few exceptions. Triple word-final consonant clusters are more numerous than quadruple and they could be morphonotactic, phonotactic or both. Triple consonant clusters mainly occur in 2nd person SG and its participle, 3rd person SG and past participle, and in Gen.SG. The expectation regarding cluster preferredness for word-final triples, i.e. morphonotactic clusters were expected to be less frequent and dispreferred was not supported by the NAD analysis. Moreover, there has been conducted a Factor analysis to check where there is a correlation between type-token frequency and NAD results.

Word-initial consonant clusters are less numerous compared to finals, as there no monoconsonantal prefixes in Standard German. Thus, in the word onset, there could be found a maximum of up to three consonants clustered together, and all of them are exclusively phonotactic. There is a moderate correlation between the degree of preferredness and the frequency in the AMC since all but one among triple clusters are preferred. Although in a word-internal position, there is a much greater variety of consonant clusters than in the peripheral positions, word-medial clusters were discussed only briefly. In the word-medial position, there are few phonotactic clusters and most morphonotactic clusters arise due to processes of compounding and affixation.

To conclude, the main finding of our research goes against with the claim that in general morphonotactic clusters are more dispreferred than phonotactic clusters (Dressler & Dziubalska-Kołaczyk 2006, Zydorowicz et al. 2016) at least for German peripheral triple consonant clusters. Another finding refers to the diachronic development of morphonotactic clusters where morphonotactic clusters have evolved into phonotactic ones through lexical development because of the loss of morpheme boundaries. In terms of typological analysis, it has been observed that word-final clusters tend to exhibit greater diversity and complexity compared to word-initial clusters. This stands in contrast to Slavic languages, Latin, Greek, and other Indo-European languages. Our analysis of the AMC corpus has revealed that this asymmetry is further evidenced by higher type and token frequencies of obstruent clusters that appear word-finally compared to those that appear word-initially. This suggests that the dominant patterns of obstruent clusters, particularly those that appear word-finally, are more likely to be used in speech and language production, resulting in a higher frequency of occurrence. As our corpus linguistic analysis shows German could be recognized as a language with a relatively high number of consonant clusters. Moreover, German displays a distinct pattern of cluster distribution. Specifically, German has a lower frequency of triple initial consonant clusters compared to Polish and Slovak. On the other hand, German has more word-final morphonotactic clusters resulting from suffixation when compared to both Slavic languages. In Polish and Slovak, the occurrence of final morphonotactic clusters is relatively limited, and these clusters are typically formed through the deletion of the word-final stem vowel in the genitive plural.

The analysis of the large Austrian Media Corpus of German allowed us to draw quantitative conclusions regarding the distribution of peripheral morphological and lexical patterns of consonant clusters. However, further research is needed to explore the distribution of consonant clusters in the word-medial position, as there may be more unique and noteworthy consonant combinations specific to the German language. All things considered, the present investigation led to the research outlined in the subsequent article, which examines the phonotactics and morphonotactics of German in comparison to another Slavic language, Russian.

1.4. Research article 3. Main differences between German and Russian (mor)phonotactics.

This study builds on previous research on contrastive studies in the domain of phonotactics and morphonotactics, namely on Polish & English (Zydorowicz 2009, Zydorowicz et al. 2016), German & Slovak (Dressler et al. 2015), Polish & German (Orzechowska & Wiese 2015). We contrast German and Russian patterns of consonantal morphonotactics vs. phonotactics from a phonological, morphological, and corpus-linguistic perspective. From a linguistic perspective, Russian, like other Slavic languages is generally considered to be more consonantal than German. In comparison to German, Russian is a more prefixing language, with a morphology that is more developed word-initially. This observation is consistent with findings on other Slavic languages, such as Polish and Slovak, which also exhibit a tendency towards a more prefix-oriented morphology. The Beats-and-Binding model of phonotactics (Dziubalska-Kołaczyk 2002, 2014) served as a standpoint for evaluating the degree of cluster preferredness. The current study has been limited to the peripheral (mor)phonotactic consonant clusters, which contain at least two obstruents since this combination is more typical for German.

Four phonemes /f/, /v/, /s/, or /z/ as the first element of word-initial Russian consonant clusters were selected for the analysis due to the following reasons. First, they occur in words not only as mono-consonantal prefixes but also in various phonotactic combinations. Therefore, they create double and triple word-initial consonant clusters both of a morphonotactic and a phonotactic nature. Second, they create voiced-voiceless pairs which, as a consequence, increase the number of possible consonant clusters. For instance, if a sibilant /s/ is followed by any voiced consonant, it becomes a voiced /z/ due to the influence of its phonological environment. Corpus research for German was based on the data from the Austrian Media Corpus. The data for the analysis of Russian consonant clusters were extracted from the "The PWN Great Russian-Polish Dictionary" (Wawrzyńczyk et al. 2007).

The Russian language allows up to four consonant clusters in the word-initial position. All analysed quadruple clusters are exclusively morphonotactic due to the concatenation of the prefix with the word coda. Triple consonant clusters are more numerous and diverse since they occur in phonotactic, morphonotactic, or ambiguous consonant clusters, which could be both. Word-finally the number of Russian consonant clusters is rather limited. Quadruple final consonant clusters behave similarly to Polish and only appear in nouns Gen. case Pl. Triple obstruent consonant clusters predominantly occur in words of foreign origin. An interesting observation for Russian, that according to the Net Auditory Distance calculator (Dziubalska-Kołaczyk 2007, 2014), the majority of triple morphonotactic clusters in word-initial position are preferred, which goes against the claim that morphonotactic clusters tend to be marked, therefore dispreferred (Dressler & Dziubalska-Kołaczyk 2006).

As it was highlighted in the abovementioned research article by Dressler & Kononenko-Szoszkiewicz (2019), unlike Russian the German language has a limited number of word-initial clusters since it doesn't tolerate mono-consonantal prefixes and there are no quadruple clusters. It could be concluded that German has a tendency towards complexity in the word-final position. Another observation for German is that word-finally there are more preferred consonant clusters, although the majority of them are morphonotactic. An important finding of this study is that, despite the type of peripheral consonant clusters, word-initially in Russian or word-finally in German, morphonotactic consonant clusters are still preferred. The conclusion could be drawn that the languages with well-developed morphology, sustain and prefer consonant clusters despite their complexity since they indicate grammatical relationships and signal meaning. Based on our research, it appears that this statement holds true at least for peripheral consonant clusters in German and Russian. However, more contrastive studies are needed to arrive at more precise results.

1.5. Research article 4. (Mor)phonotactics of Ukrainian. The study of word-initial consonant clusters.

The present study extends research on phonotactics and morphonotactics of yet another Slavic language, namely Ukrainian. This is the first study that provides a quantitative description of the consonant cluster inventory in Ukrainian. Ukrainian like other Slavic languages is considered to be a consonantal language due to a large number of consonant sounds. Isachenko (1963) posits that the presence of a phonemic contrast between plain and palatalized consonants across various articulation classes serves as evidence for the consonantal nature of the Ukrainian language. Yet the majority of studies conducted on Ukrainian phonology have been limited to providing descriptive analyses of individual sounds, and have only discussed specific groups of consonant combinations. Thus, there is a significant gap in the literature regarding the patterns of consonants, not to mention morphonotactics which has never been a focus of research in Ukrainian. Since quantitative research of Ukrainian phonotactics is rather novel, as a starting point the research of consonant clusters in the word-initial position has been selected for the following reasons. First, research on other Slavic languages, e.g., on Polish (Zydorowicz 2010, Zydorowicz et al. 2016), Russian (Dressler & Kononenko-Szoszkiewicz 2019), Slovak (Dressler et al. 2019), Croatian (Kelić & Dressler 2019) shows that word-initial position shows greater complexity in terms of the number of consonants following each other compared to word-medial or final positions. Second, a well-developed prefixation accounts for the establishment of morphonotactic consonant clusters, which are a central point of interest for comparative typological studies presented in this thesis. Lastly, the availability and accessibility of the large electronic corpus of Ukrainian allow for quantitative research and presentation of the results.

According to Zilynskyj (1979), Ukrainian tend to avoid long consonant clusters and usually create secondary syllables with a sonorant. Instead of forming a cluster, a sonorant sound is usually dropped, or a vowel is inserted between two consonants which transforms the consonant cluster into a fully voiced syllable. Following present corpus linguistic research, this statement holds since there are fewer consonant clusters in Ukrainian than in Polish and Russian. Nevertheless, word-initially there could be up to four consonants clustered together.

In this article, I have followed the same methodology as in previous papers presented in this thesis which allowed me to extend research in the field of morphonotactics and enable comparative studies within this domain. Specifically, the Beats-and-Binding syllables model was employed to assess Ukrainian consonant clusters, as proposed by Dziubalska-Kołaczyk (2002). With the help of the phonotactic Net Auditory Distance calculator, all word-initial clusters were ordered according to their preference and frequency of occurrence in the corpus (Dziubalska-Kołaczyk et al. 2007, 2009). The data for the research has been extracted from the most extensive corpus of the contemporary Ukrainian language. The General regionally annotated corpus is the biggest and, so far, the only available corpus of Ukrainian, which exhibits a great variation of reginal lexicon (Shvedova et al. 2017). Therefore, it allowed me to provide the most complete and precise quantitative analysis of type and token frequency of use of word-initial consonant clusters and gain a better understanding of the underlying patterns of this aspect of Ukrainian phonology.

The main objective of this research paper is to examine two hypotheses that have been tested on other languages discussed in this thesis before. The first hypothesis pertains to the correlation between the complexity of a cluster and its preference. The second hypothesis deals with the degree of preference and frequency of occurrence of the cluster in the corpus.

The corpus linguistic analysis revealed that double consonant clusters are the most numerous, with most of them being phonotactic. Regarding type and token frequency, the most common consonant clusters are either phonotactic or ambiguous. There are almost half as many word-initial triple consonant clusters, with the majority being morphonotactic. As expected, adding another phoneme to the word coda increases cluster complexity.

Consequently, all quadruple consonant clusters consist of a combination of the existing triple cluster and a prefix -v, indicating that all quadruple clusters are of the morphonotactic type. Both hypotheses were supported for double consonant clusters, and the first hypothesis was also verified for word-initial triples. However, the second hypothesis was only partially confirmed since most consonant clusters are dispreferred, but the five most frequent are morphonotactic and preferred.

In sum, the article revealed that Ukrainian has well-developed phonotactics. Research article 1 in this thesis demonstrated that the historical origin of the formation of morphonotactic clusters in Ukrainian is through affixation. The current study, which relied on corpus data, provided additional evidence to support this finding, especially in the context of word-initial triple and quadruple clusters. All in all, the article demonstrated a first quantitative view on the organization of Ukrainian phonemes, phonotactic preferences, and their constraints. This is the first step in the exploration of Ukrainian phonotactics, which shall be further expanded to encompass word-medial and final positions. The discussion of the comparative typological details is presented further in this thesis.

Part 2. Discussion

The second part of this PhD thesis presents the general discussion of the selected critical findings from Research article 1 (Dressler 2019 et al.), Research article 2 (Dressler & Kononenko-Szoszkiewicz, 2019), Research article 3 (Dressler & Kononenko-Szoszkiewicz 2021), Research article 4 (Kononenko-Szoszkiewicz 2023). The main goal of this section is to identify the key similarities and differences in the outcomes obtained from the selected studies and evaluate their overall implications for the research area. Additionally, any limitations or challenges encountered in the studies are identified, and potential avenues for future research are explored.

2.1. Typological characterization of the selected Germanic and Slavic languages

Maddieson's (2013) World Atlas of Language Structures presents insightful information regarding the consonant inventory of various languages. According to this atlas, German and English are classified as languages with an average number of consonants. Russian exhibits a moderately large consonant inventory, while Polish is characterized by a large consonant inventory. However, it is worth noting that the atlas does not provide any information regarding the sound inventory of the Ukrainian language.

Typically, the classification of languages into vocalic and consonantal can be determined by the number of vocalic and consonantal components in phonemic inventories or by analyzing the syllable formation and the quantity of consonant clusters (Bertinetto 2010). The variety of sound resources available in the language increases the complexity of possible sound combinations. Therefore, it is expected that more consonantal languages are supposed to have more consonant clusters which are supported by the data analysis presented in the research articles. In terms of morphological typology, there have been investigated strongly inflecting fusional languages (including Russian and Ukrainian) and a weekly inflecting language – German. In order to create a continuum in terms of the phonotactic and morphonotactic properties of the languages under investigation, the findings will be analyzed compared to earlier studies conducted on Polish and English (Zydorowicz et al. 2016).

It can be predicted that the more strongly inflecting a language is, the more morphonotactic clusters it should have. In support of this prediction, Polish has the most morphonotactic clusters among the three Slavic languages, followed by Russian and Ukrainian. For instance, if it comes to word-initial quadruple clusters based on the corpus data of Polish, there could be found 30 quadruple consonant clusters (Zydorowicz et al. 2016). Although Russian and Ukrainian share a similar set of consonant sounds, word-initially in Russian there are thirteen quadruple consonant clusters, while in Ukrainian only five.

Within the theory of (mor)phonotactics, as suggested by Dressler & Dziubalska-Kołaczyk (2006), morphonotactic clusters are supposed to be marked, and therefore dispreferred unlike phonotactic or lexical ones. When analyzing languages under the prism of morphonotactics, the question of cluster type and its preferredness has been a focal point for further typological considerations. Thus, when considering Slavic languages, it has been demonstrated that the level of morphological complexity increases as the number of segments in a cluster increase. This applies to all three languages: Polish, Russian, and Ukrainian. The prediction regarding cluster type (phonotactic or morphonotactic) and its preferredness turned out to be partially correct in that Polish morphonotactic CCs are strongly dispreferred, while lexical CCs tend to be mixed in terms of NAD. Triple CCs for Polish and Russian, in turn, were shown to be largely preferred, which goes against a theory proposed by Dressler & Dziubalska-Kołaczyk (2006) as investigated by Zydorowicz et al. (2016) and Dressler & Kononenko-Szoszkiewicz (2019). However, Ukrainian lexical double clusters are more numerous than morphonotactic and, according to the NAD calculator, are largely preferred, whereas triple morphonotactic consonant clusters are more frequent than phonotactic and strongly dispreferred. Thus, among the three Slavic languages compared, only Ukrainian seems to correspond with the theory. The claim that, in general, morphonotactic clusters are more dispreferred than phonotactic clusters has also been disproven for German peripheral triple consonant clusters but seems to be true for English (Dressler & Kononenko-Szoszkiewicz 2021, Zydorowicz et al. 2016).

An additional typological distinction can be observed in the distribution of consonant clusters in peripheral positions. German still exhibits the typical features of an inflectional language, and there are numerous complex consonant clusters in the word-final position, whereas, in English, inflection has been reduced to a minimum. Therefore, English has the least morphonotactic consonant clusters of the five languages studied. What is interesting for the typological characterization of German and English is the much greater variety and complexity of word-final than of word-initial clusters, e.g., in contrast to Slavic languages, Polish, Russian and Ukrainian. This asymmetry is also reflected in greater type and token frequencies for word-final than word-initial obstruent clusters. Type frequency asymmetries proved to be radicalized in token frequency differences, which means that the dominant patterns are more profitable. There are two reasons for this asymmetry, as mentioned in Research Article 3. First, it’s due to the outcome of prehistoric or early historic major vowel deletions in German and English word-final positions and secondly, it’s due to the optimal preservation of vowels in word-initial positions (Keyser 1975). Another distinctive characteristic of the Germanic languages is having many short derivational and inflectional suffixes, which as a rule are monoconsonantal or biconsonantal.

Phonotactic restrictions in word-peripheral positions represent another noteworthy distinction between the selected Germanic and Slavic languages. Word-initially all German triple consonant clusters start with /s/ and /ʃ/ (in English only /s/) whereas Polish, Russian, or Ukrainian allow many other consonants in this position. Word-finally, the most peripheral consonants in German are only /t, s, t͜s/, in English /t, d, s, z/. These consonants are also the preferred final consonants in double clusters. Again, there is a much greater number of final consonants that occur in the Slavic languages in the word-final clusters. Thus, it seems that in the case of strong restrictions on the selection of the most peripheral consonants, the selection is natural in the sense of not changing the (dis)preferredness of the interior consonant clusters to which they are added.

All five languages share a common feature: longer consonant clusters typically indicate the presence of a morphological boundary, while shorter clusters tend to be simpler morphologically. Therefore, the data support the statement that the longer the consonant cluster is, the more likely it is to be morphonotactic. However, in terms of morphonotactic preferredness, only Ukrainian data supported the claim that morphonotactic consonant clusters generally have a tendency to be dispreferred. This can be explained by the phonological side of the interaction between phonotactics and morphology. As presented in Research Article 4, there are two productive prefixes z- (also assimilated as s-) and v-, which give rise to the establishment of morphonotactic consonant clusters. Both in Polish and Russian, the labio-dental fricative /v/ also functions as a prefix and creates numerous morphonotactic clusters. However, in Ukrainian, the phoneme /v/ is realized as an approximant if placed between a vowel and a consonant, in the initial position before consonants and after a vowel at the end of a word, as suggested by (Pompino-Marschall et al. 2016; Buk et al., 2008; Vakulenko 2019). According to the NAD calculations, for the cluster to be preferred, it should satisfy optimal distances between the sounds. Therefore, all Ukrainian morphonotactic consonant clusters which begin with a labio-dental approximant are dispreferred because, according to the NAD formula ,the distance between C1C2 will always be greater than C2C3.

Sonority-based models have historically been employed to differentiate between well-formed and ill-formed clusters. However, contemporary approaches place greater emphasis on the frequency of occurrence and exposure as determining factors. Since the current is based on the NAD principle, one of the areas of interest is the relationship between cluster preferability and the frequency of occurrence in the corpora. One may expect that preferred clusters are supposed to be more frequent than dispreferred ones. Nevertheless, according to the corpus data provided in Research Article 2 and Research Article 3, a moderate correlation between the degree of preferredness and frequency can be observed. Similar findings have been reported in Ukrainian (Research Article 4) as well as in Polish and English (Zydorowicz et al. 2016). The overall conclusion regarding the relationship between frequency and preference is as follows: the most frequently occurring clusters consist of a mix of both dispreferred and preferred clusters.

Numerous studies have highlighted the influence of cluster frequency and phonotactic principles of sonority in psycholinguistic research. The research conducted by Wiese et al. (2017) indicates that the processing and acquisition of language in adult language users are influenced by two factors: the phonotactic principle of sonority and frequency-based input patterns. Similar results were found in the study on child-language acquisition and processing by Sommer-Lolei et al. (2021). A recent study on German by Wulfert et al. (2022) also reports that the frequency of occurrence of German consonant clusters has a significant facilitating influence on production accuracy. Similarly, Levelt, Schiller, and Levelt (2000) demonstrated that the sequence of acquiring syllable structures in French closely corresponded to the frequency of occurrence of those structures in child-directed speech. This finding, as cited by Demuth and McCullough (2009), proposes a strong connection between the frequency of syllable structures in language input and their acquisition by children.

2.2. Limitations and future research directions

This doctoral thesis provided a comprehensive overview of phonotactics and morphonotactics of the selected Slavic and Germanic languages relying on the corpus data. Such an interdisciplinary approach and the systematic analysis of large and diverse collections of language data allowed for gaining insights into language patterns and discovering their frequencies of use. However, despite computational advancements, several limitations impact the quality, generalizability, and applicability of its findings.

First, language corpora are often biased towards certain types of language use, such as written language over spoken language or formal language over informal language (see section 1.1.2). This bias can limit the generalizability of research findings. This is particularly significant in terms of phonotactics, as spoken language tends to exhibit varying production limitations and articulatory adaptation mechanisms, specifically when it comes to the realization of consonant clusters. Despite these limitations, analyzing written corpora can provide valuable insights into the phonological patterns of a language. By examining the way consonant combinations are represented in writing, one can gain a deeper understanding of the sound system of a language. This analysis can help identify common patterns and rules that govern the use of consonant clusters in written language and can also help identify irregularities or exceptions.

Furthermore, using written corpora for phonological research has the advantage of having access to large and compatible language datasets, which have been utilized in recent research to identify phonotactic and morphonotactic patterns of consonant clusters. For instance, both the AMC and GRAC corpora, which were the primary source of data for Research article 1, Research article 2, Research article 3, and Research article 4, have a similar structure and could be accessed and searched using various linguistic software tools, such as the Sketch Engine or NoSketch Engine. These tools allowed me to search for specific consonant combinations within the corpus along with keyword frequency analysis. Both corpora have a diverse and representative language, which helps to account for all possible combinations within the corpus. Moreover, employing an identical methodology to investigate both corpora of similar structures allowed me to arrive at comparable outcomes.

Reflecting on the future directions of research into phonotactics and morphonotactics, as bibliometric analysis shows (section 1.2.3), the recent decade has witnessed an increase of studies focused on language, specifically in the domains of natural language processing (NLP), speech recognition and language modeling. This bibliometric analysis reinforces the necessity for a deeper and more systematic exploration of phonotactic patterns, particularly emphasizing the use of spoken language corpora. Consonants play a vital role by providing essential cues for accurate recognition and understanding of spoken language, as they carry phonemic distinctions that differentiate words and convey meaning, making them indispensable for precise transcription and comprehension. Therefore, studying and analysing phonotactic patterns, constraints, and combinatory possibilities of consonants is essential for developing accurate language models.

Conclusion

The present PhD project aimed to investigate and analyze the phonotactic and morphonotactic patterns found in selected Germanic and Slavic languages, utilizing corpus data as the primary source. In order to examine the consonant clusters within these languages, I adopted an interdisciplinary perspective, employing a quantitative approach to explore phonotactic preferences. The utilization of written corpus data was crucial, as it provided tangible evidence regarding the presence, variability, language-specific constraints, and frequency of occurrence of consonant clusters across different word positions.

Moreover, a bibliometric analysis of the Scopus database presented an overview of the current state of the art in linguistic studies that employ interdisciplinary approaches and corpus linguistic methods. A discussion has been shown on the prominent trends and prevailing types of corpora used in corpus linguistic analysis while also providing information on the dominant languages explored in such studies. The primary objective of conducting this bibliometric analysis was to determine the position of phonetic and phonological research within the realm of corpus linguistics.

The findings of the bibliometric analysis highlighted a substantial number of studies conducted in the past decade, mainly focusing on language sounds within the domains of natural language processing, language perception and acquisition, and machine learning. Consequently, studying phonotactic patterns, relying on vast amounts of language data opens up new possibilities for exploring the mechanisms that shape our understanding of language.

Overall, my doctoral thesis is a compilation of four thematically related research articles which investigate and compare phonotactic and morphonotactic patterns of German, Russian, and Ukrainian with reference to the existing research within this domain, primarily on Polish and English by (Zydorowicz et al. 2016). The choice of languages is motivated by the fact that selected languages belong to different language families and present substantial differences concerning phonological and morphological patterns. Moreover, all four articles follow the same methodology for the exploration of the preferences of consonant clusters and are evaluated according to the Net Auditory Distance principle (Dziubalska-Kołaczyk, 2014). The analysis of consonant clusters, categorized as either phonotactic or morphonotactic, raises various research questions, several of which were addressed in the research articles included in the doctoral thesis. Specifically, what are the type and token frequencies of specific consonant clusters in the corpora, their preference status based on type, and is there a correlation between cluster preference and frequency?

Research Article 1 (Dressler et al. 2019) is an introductory and explanatory piece that explores the emergence of morphonotactic consonant clusters from a diachronic standpoint. By analyzing data from selected Slavic, Germanic, Baltic, and Romance languages, the authors aimed to identify major historical processes that contributed to the evolution of morphonotactic consonant clusters. The study concludes that these processes involve phenomena such as vowel loss, Indo-European influence, affixation, compounding, metathesis, final and consonant epenthesis. However, most morphonotactic clusters can be categorized into two main types: phonologically derived clusters resulting from vowel loss, observed in Slavic languages, and morphologically derived clusters resulting from concatenation, observed in Germanic languages.

The subsequent two articles, Research Article 2 and Research Article 3 provided a quantitative analysis of German consonant clusters utilizing data from the extensive Austrian Media Corpus. For the first time, these articles presented a comprehensive characterization of German patterns of consonantal morphonotactics and phonotactics from various perspectives, including phonological, morphological, typological, and corpus linguistic viewpoints. Specifically, word-initial and word-final consonant clusters are thoroughly analyzed based on phonotactic preferences established within the framework of the Beats-and-Binding Model (Dziubalska-Kołaczyk 2002). German data was compared to the Russian data which allowed to arrive at first tentative generalisations regarding typological differences. One of the main findings is that claim that phonotactic clusters are more preferred than morphonotactic clusters has been disproven. However, the analysis of the Ukrainian as presented in Research Article 4, confirmed a general presumption that morphonotactic clusters tend to be marked and therefore dispreferred. Furthermore, no significant correlation was observed between cluster preference and its frequency across all languages examined. However, based on the analysis of five languages, it can be inferred that the longer cluster is, the greater probability it to be morphonotactic.

Summing up, the findings presented in this doctoral thesis provide valuable insights into the study of phonotactic and morphonotactic patterns by application of cross-linguistic methods. These findings contribute to the existing body of knowledge in the field by introducing new perspectives and shedding light on previously unexplored aspects of language patterns. By adopting a comparative approach across different languages, this research expands our understanding of the underlying principles and mechanisms that govern phonotactic and morphonotactic phenomena.

Abstract

This PhD thesis aimed to investigate and analyse the phonotactic and morphonotactic patterns which occur in German, English, Polish, Ukrainian and Russian. In order to examine the consonant clusters within these languages, I adopted an interdisciplinary perspective by utilizing corpus data as the primary source to explore phonotactic preferences. The utilization of written corpus data allowed to investigate variety, language-specific constraints, and frequency of occurrence of consonant clusters in different word positions of the studied languages. This crucial information was then applied to draw comparative generalizations.

Furthermore, to fulfill the objectives of this thesis, a comprehensive bibliometric analysis was carried out on the Scopus database. This analysis provides a comprehensive overview of the current landscape of linguistic studies utilizing corpus linguistic methods to examine phonetic and phonological phenomena. In this context, I discuss the prominent trends and prevailing types of corpora used in corpus linguistic analysis (written vs. spoken), while also providing information on the dominant languages explored in such studies. The primary objective of the bibliometric analysis was to determine the position of phonetic and phonological research within the realm of corpus linguistics. The findings of the bibliometric analysis highlighted a substantial number of studies conducted in the past decade, particularly focusing on the study of sounds within the domains of natural language processing, language perception and acquisition, and machine learning.

Overall, my doctoral thesis is a compilation of four thematically related research articles which investigate and compare phonotactic and morphonotactic patterns of German, Russian, and Ukrainian with reference to the existing research within this domain, primarily on Polish and English (Zydorowicz et al. 2016). All four articles follow the same methodology for exploring phonotactic preferences of consonant clusters and apply the Net Auditory Distance principle (Dziubalska-Kołaczyk, 2014). The analysis of consonant clusters, categorized as either phonotactic or morphonotactic, raises various questions, a few of which were addressed in the research articles included in the doctoral thesis. Namely, what is the type and token frequency of specific consonant clusters present in the corpora of the selected languages, what is their preference status according to

the NAD, and whether is there a relationship between cluster preference and its frequency?

Research Article 1 is an introductory and explanatory study exploring the emergence of morphonotactic consonant clusters from a diachronic standpoint. By analysing data from selected Slavic, Germanic, Baltic, and Romance languages and others, we aimed to identify major historical processes that contributed to the evolution of morphonotactic consonant clusters. The study concludes that these processes involve the following: vowel loss, Indo-European ablaut, affixation, compounding, metathesis, final and consonant epenthesis. Moreover, we concluded that most morphonotactic clusters could be categorized into two main types: phonologically derived clusters resulting from vowel loss, as observed in Slavic languages, and morphologically derived clusters resulting from concatenation, as observed in Germanic languages.

The subsequent two articles, Research Article 2 and Research Article 3, provided a quantitative analysis of German consonant clusters utilizing data from the Austrian Media Corpus. These articles provided a quantitative characterization of German patterns of consonantal morphonotactics and phonotactics from various perspectives, including phonological, morphological, typological, and corpus linguistics. Specifically, word-initial and word-final consonant clusters are analysed within the framework of the Beats-and-Binding Model (Dziubalska-Kołaczyk 2002). German data was compared to the Russian data which allowed us to arrive at first tentative generalisations regarding typological differences. One of the main findings of these studies is that the claim that phonotactic clusters are more preferred than morphonotactic clusters has been disproven. However, the analysis of the Ukrainian as presented in Research Article 4, confirmed a general presumption that morphonotactic clusters tend to be marked and therefore dispreferred. Furthermore, no significant correlation was observed between cluster preference and its frequency across all languages examined. However, based on the analysis of five languages, it can be inferred that the longer cluster is, the more likely it is to be morphonotactic.

To conclude, the findings presented in this doctoral thesis provide insights into the study of phonotactic and morphonotactic patterns by application of cross-linguistic methods of research. By adopting a comparative approach across different languages, this research expands our understanding of the underlying principles and mechanisms that govern phonotactics and morphonotactics.

Streszczenie

Niniejsza rozprawa doktorska prezentuje wyniki badań w zakresie fonologii, w szczególności fonotaktyki i morfonotaktyki. Celem pracy było zbadanie i przeanalizowanie zbitek fonotaktycznych i morfonotaktycznych występujących w językach niemieckim, angielskim, polskim, ukraińskim i rosyjskim. W celu zbadania zbitek spółgłoskowych w tych językach, przyjęłam interdyscyplinarną perspektywę, wykorzystując dane korpusowe jako podstawowe źródło do zbadania preferencji fonotaktycznych. Wykorzystanie pisemnych danych korpusowych pozwoliło na zbadanie różnorodności występujących zbitek spółgłoskowych, ograniczeń charakterystycznych dla danego języka, a także częstotliwości występowania takich zbitek an początku, w środku, lub na końcu słowa.

Ponadto, na potrzeby niniejszej rozprawy przeprowadziłam analizę bibliometryczną bazy publikacji naukowych Scopus, przedstawiając przegląd aktualnego stanu wiedzy w badaniach językoznawczych, które wykorzystują metody językoznastwa korpusowego do badania zjawisk fonetycznych i fonologicznych. W tym kontekście omawiam w rozprawie główne trendy oraz najczęściej występujące typy korpusów (m.in. języka pisanego, języka mówionego) wykorzystywanych w analizie językowej korpusów, a także wskazuję dominujące języki będące przedmiotem badań. Głównym celem przeprowadzenia analizy bibliometrycznej było określenie rozległości tematycznej publikacji wykorzystujących językoznastwo korpusowe do badania fonetyki i fonologii. Wyniki analizy bibliometrycznej wskazały na znaczną liczbę badań przeprowadzonych w ciągu ostatniej dekady, w szczególności koncentrujących się na badaniu dźwięków w domenach przetwarzania języka naturalnego, percepcji i akwizycji języka oraz uczenia maszynowego.

Niniejsza rozprawa doktorska jest spójnym tematycznie zbiorem czterech artykułów, dotyczących zbitek fonotaktycznych i morfonotaktycznych języków niemieckiego, rosyjskiego i ukraińskiego, a także analizy wyników tych badań w odniesieniu do zależności występujących w językach polskim i angielskim (Zydorowicz et al. 2016). Powyższy wybór języków jest motywowany faktem, że należą one do różnych rodzin językowych, ale zarazem wykazują znaczne różnice w zakresie cech fonologicznych i morfologicznych. We wszystkich czterech artykułach zastosowałam tę samą metodologię badania preferencji fonotaktycznych zbitek spółgłoskowych i oceniłam ich preferencję zgodnie z *Net Auditory Distance* (NAD) *Principle* (Dziubalska-Kołaczyk 2014). W rozprawie doktorskiej skupiłam się przede wszystkim na analizie typów i częstotliwości występowania specyficznych zbitek spółgłoskowych obecnych w korpusach wybranych języków, ich statusie preferencyjnym według NAD oraz badaniu związku zachodzącego pomiędzy częstotliwością zbitek a ich preferencją.

Artykuł Badawczy 1 prezentuje wprowadzenie do tematyki morfonotaktyki, ze szczególnym uwzględnieniem mechanizmu powstawania morfonotaktycznych zbitek spółgłoskowych. Analizując dane z wybranych języków słowiańskich, germańskich, bałtyckich, romańskich i innych, zidentyfikowaliśmy wsytąpienie procesów takich jak: utrata samogłosek, indoeuropejska apofonia, afiksacja, tworzenie wyrazów złożonych (ang. *compounding*), metateza, epenteza końcowa i spółgłoskowa. Stwierdziliśmy, że większość zbitek morfonotaktycznych można podzielić na dwa główne typy: zbitki powstałe fonologicznie w wyniku utraty samogłosek (języki słowiańskie), oraz zbitki powstałe morfologicznie w wyniku łączenia (ang. *concatenation*) (języki germańskie).

Kolejne dwa artykuły, Artykuł Badawczy 2 i Artykuł Badawczy 3 przedstawiają analizę ilościową zbitek spółgłoskowych występujących w języku niemieckim, wykorzystując dane z Austriackiego Korpusu Medialnego (*Austrian Media Corpus*). Publikacje opisują charakterystykę cech morfonotaktycznych i fonotaktycznych zbitek spółgłoskowych z różnych perspektyw, w tym fonologicznej, morfologicznej, typologicznej i językoznawczej. Początkowe i końcowe zbitki spółgłoskowe zostały przeanalizowane w ramach modelu *Beats-and-Binding* (Dziubalska-Kołaczyk 2002). Dane dla języka niemieckiego zostały porównane z danymi dla języka rosyjskiego, co pozwoliło na dokonanie wstępnych uogólnień w zakresie różnic typologicznych. Jednym z głównych wniosków płynących z badania jest negatywna weryfikacja hipotezy, zgodnie z którą zbitki fonotaktyczne miałyby być preferowane ponad zbitki morfonotaktyczne (zgodnie z stopniami preferencji NAD). Jednocześnie analiza danych dla języka ukraińskiego przedstawiona w Artykule Badawczym nr 4 potwierdziła ogólne założenie, że zbitki morfonotaktyczne mają tendencję do bycia nacechowanymi (ang. *marked*), a zatem nie są preferowane (ang. *dispreferred*). Ponadto w żadnym z badanych języków nie zaobserwowałam istotnej korelacji między stopniem preferencji zbitek a częstotliwością ich występowania. Na podstawie analizy wybranych pięciu języków można natomiast wywnioskować, że im dłuższa jest zbitka, tym większe prawdopodobieństwo, że będzie ona morfonotaktyczna.

Podsumowując, wyniki przedstawione w niniejszej rozprawie doktorskiej stanowią istotny wkład w badania porównawcze nad fonotaktycznymi i morfonotaktycznymi cechami języków słowiańskich i germańskich.

References

Baumann, Andreas and Kamil Kaźmierski. 2016. “A dynamical-systems approach to the evolution of morphonotactic and lexical consonant clusters in English and Polish”, *Yearbook of the Poznan Linguistic Meeting*, 1: 115-139.

Bertinetto, Pier Marco. 2010. “Sillaba”, in: Raffaele Simone (cur., con la collaboraz. di Gaetano Berruto e Paolo d'Achille), *Enciclopedia dell’Italiano*. Roma, Istituto dell’Enciclopedia Italiana [vers. preliminare in Quaderni del Laboratorio di Linguistica della SNS, 1-9].

Biber, Douglas and Randi Reppen. 2015. *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press.

Biber, Douglas, Susan Conrad, and Randi Reppen. 1994. “Corpus-based approaches to issues in applied linguistics”. *Applied Linguistics*, 15: 169–187.

Blevins, Juliette. 2007. “The importance of typology in explaining recurrent sound patterns”, *Linguistic Typology* 11: 107–113.

Buk, Solomija, Ján Mačutek and Andrij Rovenchak, 2008. “Some properties of the Ukrainian writing system”. *Glottometrics* 16: 63–79.

Calderone, Basilio, Chiara Celata, Katharina Korecky-Kröll and Wolfgang U. Dressler 2014. “A computational approach to morphonotactics: evidences from German”. *Language Sciences* 46: 59-70.

Chomsky, Noam and Morris Halle. 1965. “Some controversial questions in phonological theory”. *Journal of Linguistics* 1: 97–138.

Cori, Marcel and Sophie David. 2008. “Les corpus fondent-ils une nouvelle linguistique?”, *Languages* 171: 111 -129.

Davies, Mark. 2008- . *The Corpus of Contemporary American English (COCA)*. Available online at https://www.english-corpora.org/coca/.

Delais-Roussarie, Elisabeth. and Hiyon Yoo. 2014. “Corpus and Research in Phonetics and Phonology: Methodological and Formal Considerations”, in: Jacques Durand, Ulrike Gut, and Gjert Kristoffersen (eds), *The Oxford Handbook of Corpus Phonology*, 193-213.

Demuth, Katherine and Elizabeth McCullough. 2009. “The longitudinal development of clusters in French”, *Journal of Child Language* 36, 425–448.

Dressler, Wolfgang U. and Katarzyna Dziubalska-Kołaczyk. 2006. “Proposing morphonotactics”, *Italian Journal of Linguistics* 18: 249–266.

Dressler, Wolfgang U. 1996. “On a semiotic theory of preferences in language”, in: Michael Haley and Michael Shapiro (eds.), *The Peirce Seminar Papers. Essays in Semiotic Analysis. Proceedings of the International Colloquium on Language and Peircean Sign Theory*, New York, Berghahn Books: 389-415.

Dressler, Wolfgang U., Katarzyna Dziubalska-Kolaczyk and Lina Pestal. 2010. “Change and variation in morphonotactics”, *Folia Linguistica* 31: 51–67.

Dressler, Wolfgang U., Miroslava Hliničanová, Matej Ďurčo, Karlheinz Mörth and Katharina Korecky-Kröll. 2015. “Phonotactic vs. morphonotactic obstruent clusters in Slovak and German”. *Italian Journal of Linguistics* 27, 1: 45-60.

Dressler, Wolfgang. U. 1985. *Morphonology: The dynamics of derivation*. Ann Arbor: Karoma

Durand, Jacques, 2009. “On the scope of linguistics: Data, intuitions, corpora”, in: Yuji Kawaguchi, Makoto Minegishi, and Jacques Durand (eds), *Corpus Analysis and Variation in Linguistics* 6: 25-52.

Durand, Jacques, Ulrike Gut and Gjert Kristoffersen. 2013. “Introduction”. *Oxford Handbooks Online.*

Dziubalska-Kołaczyk, Katarzyna. 2014. “Explaining phonotactics using NAD”. *Language Sciences* 46: 6–17.

Dziubalska-Kołaczyk, Katarzyna. 2002. *Beats-and-Binding Phonology*. Frankfurt am Main: Peter Lang.

Dziubalska-Kołaczyk, Katarzyna. 2009. “NP extension: B&B phonotactics”, *Poznań Studies in Contemporary Linguistics* 45, 1: 55-71.

Dziubalska-Kołaczyk, Katarzyna and Grzegorz Krynicki, 2007. Phonotactic Preferences in Polish, English and German: Quantitative Perspective. (Paper presented at the 38th Poznan Linguistic Meeting , Poznan, 13–16 September 2007.).

Dziubalska-Kołaczyk, Katarzyna, Dawid Pietrala, Grzegorz Aperliński. 2014. “The NAD Phonotactic Calculator: An Online Tool to Calculate Cluster Preference in English, Polish and Other Languages”, [On:] http://wa.amu.edu.pl/wa/files/The_NAD_Phonotactic_Calculator.pdf.

Freiberger, Eva. 2007. “Morphonotaktik im Erstspracherwerb” [Morphonotactics in first language acquisition], *Wiener Linguistische Gazette* 74: 1-23.

Frisch, Stefan. 2012. “Phonotactic patterns in lexical corpora”, in: Abigail C. Cohn, Cécile Fougeron, and Marie K. Huffman (eds.), *The Oxford Handbook of Laboratory Phonology*, Oxford University Press, 458-470.

Hicks, Diana. 2004. “The four literatures of social science”, in: Henk F. Moed, Wolfgang Glänzel, and Ulrich Schmoch (eds.) *Handbook of Quantitative Science and Technology Research*. Kluwer Academic Publishers, 473–496.

Hilary Arksey & Lisa O’Malley. 2005. “Scoping studies: towards a methodological framework”, *International Journal of Social Research Methodology* 8, 1: 19-32.

Hill, Archibald A. 1958. *Introduction to linguistic structures: From sound to sentence in English*. New York.

Hyman, Larry. 2007. “Where’s Phonology in Typology?”, *Linguistic Typology* 11, 265–271.

Hyman, Larry M. and Frans Plank (eds.) 2018. “Phonological typology. (Phonology and Phonetics 23)”, *Phonology* 36, 3: 537-542.

Isachenko, Aleksandr. 1963. Opyt tipologicheskogo analiza slavjanskikh jazykov [Experience of typological analysis of Slavic languages]. *Novoe v lingvistike* 3, 106–121.

Kelić, Maja and Wolfgang U. Dressler. 2019. “The development of morphonotactic and phonotactic word-initial consonant clusters in Croatian first-language acquisition”. *Suvremena Lingvistika* 45, 2: 179-200.

Keyser, Samuel Jay. 1975. “Metathesis and Old English Phonology,” *University of Massachusetts Occasional Papers in Linguistics* 1, 1:7.

Kendall, Tyler. 2013. “Speech Rate, Pause and Sociolinguistic Variation: Studies in Corpus Sociophonetics”. Springer.

Levelt, Clara C, Niels O. Schiller and Willem J. Levelt. 2000. “The acquisition of syllable types”, *Language Acquisition* 8, 237–264.

Maddieson, Ian. 2013. “Consonant Inventories”. In: Dryer, Matthew S. and Haspelmath, Martin (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/), (date of access 11 March 2023).

Matviias, Ivan. (ed.) 1988-2001. Atlas ukrainskoi movy. [Atlas of the Ukrainian Language in Three Volumes].

Martin, Ben, Puay Tang, Molly Morgan, Wolfgang Glänzel, Stefan Hornbostel, Gerhard Lauer, Gerard Lenclud, Luisa Lima, Charles Oppenheim, Peter van den Besselaar and Milena Zic-Fuchs. 2010. *Towards a bibliometric database for the social sciences and humanities—A European scoping project: A report produced for DFG, ESRC, AHRC, NWO, ANR and ESF.* Sussex: Science and Technology Policy Research Unit.

Meyer, Antje S., Falk Huettig and Willem J.M. Levelt. 2016. "Same, different, or closely related: What is the relationship between language production and comprehension?", *Journal of Memory and Language* 89: 1-7.

Munn, Zachary, Micah D. J. Peters, Cindy Stern, Catalin Tufanaru, Alexa McArthur & Edoardo Aromataris. 2018. "Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach". *BMC Med Res Methodology* 18, 143.

Nederhof, Anton J. 2006. "Bibliometric monitoring of research performance in the Social Sciences and the Humanities: A Review", *Scientometrics* 66: 81–100.

Orzechowska, Paula and Richard Wiese. 2015. "Preferences and variation in word-initial phonotactics: A multi-dimensional evaluation of German and Polish", *Folia Linguistica* 49: 439 - 486.

Pierrehumbert, Janet B. 2003. "Phonetic Diversity, Statistical Learning, and Acquisition of Phonology". *Language and Speech*, 46(2–3), 115–154.

Pompino-Marschall, Bernd, Steripolo, Elena and Marzena Żygis. 2016. "Ukrainian". *Journal of the International Phonetic Association* 47, 1–9.

Prince, Alan and Paul Smolensky. 1993. *Optimality Theory: Constraint interaction in generative grammar*. New Brunswick, NJ: Rutgers University Center for Cognitive Science

Sivertsen, Gunnar. 2014. "Scholarly publication patterns in the social sciences and humanities and their coverage in Scopus and Web of Science". in: *Proceedings of the science and technology indicators conference*, Leiden: Universiteit Leiden – CWTS, 598–604.

Sommer-Lolei, Sabine, Katharina Korecky-Kröll, Markus Christiner and Wolfgang U. Dressler. 2021. "The acquisition and processing of (mor)phonotactic consonant clusters in German", in: Wolfgang U. Dressler, Basilio Calderone, Sabine Sommer-Lolei, Katharina Korecky- Kröll (eds.), *Experimental, Acquisitional and*

Corpus linguistic Approaches to the Study of Morphonotactics. Vienna: Verlag der Österreichischen Akademie der Wissenschaften, 77-101.

Trubetskoy, Nikolay. 1931. "Gedanken über Morphonology". *Travaux de Cercle Linguistique de Prague*, University of Michigan Press, 4: 53–61.

Ransmayr, Jutta. 2018. "Austrian Media Corpus", in: Börner, Ingo, Wolfgang Straub, Christian Zolles, Wien (eds.), *Germanistik digital. Digital Humanities in der Sprach- und Literaturwissenschaft*: Facultas, 168–182.

Resch, Claudia and Wolfgang U. Dressler. 2017. *Digitale Methoden der Korpusforschung in Österreich*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Riitta, Välimaa-Blum. 2005. "Cognitive Phonology in Construction Grammar: Analytic Tools for Students of English", Berlin: De Gruyter Mouton.

van Eck Nees Jan and Ludo Waltman. 2010. "Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping", *Scientometrics*, 84: 523–38.

Vakulenko, Maksym O. 2019. "Ukrainian Consonant Phones in the IPA Context with Special Reference to /v/ and /gh/." In: *Linguistica online* 22: 36-61.

Voormann, Holger and Ulrike Gut. 2008. "Agile corpus creation", *Corpus Linguistics and Linguistic Theory* 4, 2: 235-251.

Wawrzyńczyk, Jan, Magdalena Kuratczyk, Eliza Małek and Halina Bartwicka. 2007. *Wielki słownik rosyjsko–polski*. Warsaw: PWN.

Wiese, Richard, Paula Orzechowska, Phillip M. Alday and Christiane Ulbrich. 2017. "Structural Principles or Frequency of Use? An ERP Experiment on the Learnability of Consonant Clusters". *Frontiers in psychology*, 7: 2005.

Wulfert, Sophia, Peter Auer and Adriana Hanulíková. 2022. "Speech Errors in the Production of Initial Consonant Clusters: The Roles of Frequency and Sonority." *Journal of speech, language, and hearing research* 10: 3709-3729.

Zydorowicz, Paulina, Paula Orzechowska, Michał Jankowski, Katarzyna Dziubalska-Kołaczyk, Piotr Wierzchoń and Dawid Pietrala. 2016. *Phonotactics and morphonotactics of Polish and English: Theory, description, tools and applications.* Adam Mickiewicz University Press.

Zydorowicz, Paulina. 2010. "Consonant Clusters Across Morpheme Boundaries: Polish Morphonotactic Inventory and its Acquisition", *Poznań Studies in Contemporary Linguistics* 46, 4: 565-588.

Zydorowicz, Paulina. 2009. “English and Polish morphonotactics in first language acquisition”. [Unpublished PhD dissertation, Adam Mickiewicz University, Poznań.].

Zydorowicz, Paulina. 2019. *Polish (mor)phonotactics in first language acquisition, connected speech and cluster processing*. Wydawnictwo Naukowe UAM.

Zydorowicz, Paulina. 2020. “The Acquisition of Polish Phonotactics at Word Edges: A Markedness Account”, in: Elena Babatsouli, Bristol (eds.), *Under-reported Monolingual Child Phonology*, Blue Ridge Summit: Multilingual Matters: 145-176.

Appendix 1. Search query to retrieve data from Scopus

TITLE-ABS-KEY ((written OR spoken) AND (corpus OR corpora)) AND (LIMIT-TO (DOCTYPE, “ar”) OR LIMIT-TO (DOCTYPE, “ch”) OR LIMIT-TO (DOCTYPE, “bk”)) AND (LIMIT-TO (PUBYEAR,2020) OR LIMIT-TO (PUBYEAR,2019) OR LIMIT-TO (PUBYEAR,2018) OR LIMIT-TO (PUBYEAR,2017)) OR LIMIT-TO (PUBYEAR,2016)) OR LIMIT-TO (PUBYEAR,2015)) OR LIMIT-TO (PUBYEAR,2014)) OR LIMIT-TO (PUBYEAR,2013)) OR LIMIT-TO (PUBYEAR,2012)) OR LIMIT-TO (PUBYEAR,2011)) OR LIMIT-TO (PUBYEAR,2010) OR LIMIT-TO (PUBYEAR,final))

Appendix 2. List of languages as an object of corpus research

1. English: 505
2. Spanish: 164
3. British English: 85
4. French: 84
5. American English: 73
6. German: 71
7. Chinese: 43
8. Russian: 36
9. Arabic: 34
10. Dutch: 33
11. Italian: 31
12. Portuguese: 28
13. Estonian: 27
14. Czech: 26
15. Catalan: 20
16. Japanese: 18
17. Polish: 16
18. Australian English: 15
19. Mandarin Chinese: 19
20. Swedish: 14
21. Danish: 14
22. Persian: 14
23. Greek: 13
24. Afrikaans: 12
25. Finnish: 12
26. Hebrew: 11
27. Croatian: 11
28. Norwegian: 10
29. Brazilian Portuguese: 10
30. Lithuanian: 9
31. Korean: 8
32. Basque: 8
33. Romanian: 8
34. Taiwanese Mandarin: 7
35. Urdu: 7
36. Slovene: 7
37. Malay: 6
38. Thai: 6
39. Indonesian: 6
40. Turkish: 6
41. Asian English: 6
42. Irish: 6
43. Hungarian: 6
44. Serbian: 6
45. Indian English: 5
46. Hindi: 5
47. New Zealand English: 5
48. Latin: 5
49. Icelandic: 5
50. Hong Kong English: 4
51. Bengali: 4
52. Tuvan: 4
53. Slovak: 4
54. African English: 4
55. Old Croatian: 4
56. American Sign: 4
57. Welsh: 3
58. Canadian English: 3
59. Galician: 3
60. Irish English: 3
61. Philippine English: 3
62. Kalmyk: 3

63. Faroese: 3
64. Pakistani English: 3
65. Arabic English: 2
66. Singapore English: 2
67. Flemish: 2
68. Indian: 2
69. Mexican Spanish: 2
70. Agul: 2
71. Palestinian Arabic: 2
72. Ladin: 2
73. Ukrainian: 2
74. Vietnamese: 2
75. American Norwegian: 2
76. Australian Sign: 2
77. Castilian: 2
78. Brazilian: 2
79. Cantonese: 2
80. Austrian German: 2
81. Egyptian: 2
82. Taiwanese: 2
83. Chilean Spanish: 2
84. Nazarene: 2
85. Zulu: 2
86. African American English: 3
87. Scottish: 2

Languages have been only mentioned once: Malayalam, Sicilian, Arabic Sign, Armenian, American Danish, English, Arabizi, Punjabi, Kreol, West African, Russian Sign, White South African English, Belgian, Québécois French, Korean(Ized) English, Uzbek, Jordanian Arabic, Tunisian, Papuan, Indonesian English, Urdu Sindhi, Kazakh, Ruruuli-Lunyala, Toba, Balochi, Swiss German, Trentino, Cimbrian, Tyrolean, Burmese, Argentine Spanish, Indonesian , Malayan, Namibian English, Papiamento, Lycian, Late Egyptian, Francoprovençal, West Flemish, French Flemish, Italian Sign, Marathi, Slovenian, South African English, Tehuelche, Galician Spanish, Slavonic, Old Hungarian, Israeli Hebrew, Romani, Vurës, Romangol, Greek Sign, Old Swedish, Dutch Sign, Old English, Nigerian English, Catalan Sign , Chinese Pidgin Spanish, Albanian, Bangla, Sign Auslan, Aramaic, Bantu, Luxembourgish, Korelian, Guarani, New High German, Madurese, Swedish Sign, Karelian, Andalusian Spanish, Francoprovencąl, Trinidadian English, Latvian, Quebec French, Cameroon Pidgin English, Punjubi, Nigerian Arabic, Gaelic, Swedish, Flemish Dutch, Tunisian Arabic, Iraqi English, Glagolitic, Argentine Danish, Flemish , Sardinian, British Sign, Chendungun, Polish English, Japanese English, Palestenian, Hittite, Barwe, Telugu, Afrikaans, Nheengatú, French English, Xhosa, Panjabi, Argentine Sign, Tamil, German Sign, Swiss German Sign, African American Sign, Sign Language of The Netherlands, Malaysian, Kiranti, Korean English, Spanish Romani, Lingala, Northern Sotho, Swabian, Paini, Mixtec, Tibetan, Luwian, Turkish Sign, Acadian French, Macedonian, Uyghur, Bahamian Standard English, Canadian French, Cantonese.

Appendix 3. Author contribution statement: Dressler et al. (2019)

Poznań, 1st June 23

Author contribution statement

Dressler, Wolfgang U., Alona Kononenko, Sabine Sommer-Lolei, Katharina Korecky-Kröll, Paulina Zydorowicz & Laura Kamandulytė-Merfeldienė. 2019. "Morphological richness, transparency and the evolution of morphonotactic patterns", *Folia Linguistica Historica, Folia Linguistica* 40(1), 85-106.

All co-authors of the research article titled: "Morphological richness, transparency and the evolution of morphonotactic patterns", published in Folia Liguistica Historica in July 2019, hereby declare that they have contributed to the research article in the following way:

- emer. o. Univ.-Prof. Mag. Dr. U. Wolfgang Dressler (25%): study conception and design, structure planning, some manuscript writing; supervision;
- Mgr Alona Kononenko (15%): data analysis (Ukrainian & Russian); investigation, draft manuscript preparation;
- Mgr. Sabine Sommer-Lolei (15%): investigation, data analysis (German); method description, draft manuscript preparation; review and editing;
- Dr. Katharina Korecky-Kröll (15%): investigation, data analysis (German); draft manuscript preparation;
- prof. UAM dr hab. Paulina Zydorowicz (15%): investigation, data analysis (Polish); draft manuscript preparation;
- Dr. Assoc. Prof. Laura Kamandulytė-Merfeldienė (15%): investigation; data analysis (Baltic languages); draft manuscript preparation;

Authors:

emer. o. Univ.-Prof. Mag. Dr. U. W. Dressler

Dr. Katharina Korecky-Kröll

mgr. Alona Kononenko

prof. UAM dr hab. Paulina Zydorowicz

mgr. Sabine Sommer-Lolei

Dr. Assoc. Prof. Laura Kamandulytė-Merfeldienė

PhD supervisors:

prof. dr hab. Katarzyna Dziubalska – Kołaczyk

emer. o. Univ.-Prof. Mag. Dr. U. Wolfgang Dressler

Appendix 4. Author contribution statement: Dressler & Kononenko-Szoszkiewicz (2019)

Poznań, 1st June 23

Author contribution statement

Dressler, Wolfgang U. and Alona Kononenko-Szoszkiewicz. 2019. "Main Differences Between German and Russian (Mor)phonotactics", In.: Wrembel, M., Kiełkiewicz-Janowiak, A., & Gąsiorowski, P. (eds.) *Approaches to the Study of Sound Structure and Speech: Interdisciplinary Work in Honour of Katarzyna Dziubalska-Kołaczyk (1st ed.)*. Routledge.

All co-authors of the research article titled: "Main Differences Between German and Russian (Mor)phonotactics", published in Routledge November 2019, hereby declare that they have contributed to the research article in the following way:

- W.U. Dressler (40%): study conception and design, structure planning, some manuscript writing; supervision
- Mgr. Alona Kononenko-Szoszkiewicz (60%): data collection and analysis, manuscript writing, reviewing and editing

Authors:

....

emer. o. Univ.-Prof. Mag. Dr. U. Wolfgang Dressler

PhD supervisors:

....

prof. dr hab. Katarzyna Dziubalska – Kołaczyk

....

Mgr Alona Kononenko-Szoszkiewicz

....

emer. o. Univ.-Prof. Mag. Dr. U. Wolfgang Dressler

Appendix 5. Author contribution statement: Dressler & Kononenko-Szoszkiewicz (2021)

Poznań, 1st June 23

Author contribution statement

Dressler, Wolfgang U. and Alona Kononenko-Szoszkiewicz, 2021, "German phonotactic vs. morphonotactic obstruent clusters: a corpus linguistic analysis ", In.: W. U. Dressler, B. Calderone, S. Sommer-Lolei, K. Korecky-Kröll (eds.) *Experimental, Acquisitional and Corpus linguistic Approaches to the Study of Morphonotactics*, Austrian Academy of Sciences Press.

All co-authors of the research article titled: "German phonotactic vs. morphonotactic obstruent clusters: a corpus linguistic analysis "published in Austrian Academy of Sciences Press, 2021, hereby declare that they have contributed to the research article in the following way:

- Wolfgang U. Dressler (35%): study conception and design, structure planning, some manuscript writing; supervision
- Mgr. Alona Kononenko-Szoszkiewicz (65%): data collection and analysis (incl. statistics), manuscript writing, reviewing and editing

Authors:

emer. o. Univ.-Prof. Mag. Dr. U. Wolfgang Dressler

Alona Kononenko-Szoszkiewicz

Mgr Alona Kononenko-Szoszkiewicz

PhD supervisors:

prof. dr hab. Katarzyna Dziubalska – Kołaczyk

emer. o. Univ.-Prof. Mag. Dr. U. Wolfgang Dressler

Appendix 6. Author contribution statement: Kononenko-Szoszkiewicz (2023)

Poznań, 1st June 23

Author contribution statement

Alona Kononenko-Szoszkiewicz, 2023, "(Mor)phonotactics of Ukrainian. The study of word-initial consonant clusters ", [Manuscript accepted for publication].

The author of the research article titled: "(Mor)phonotactics of Ukrainian. The study of word-initial consonant clusters ", hereby declare that she contributed to the research article in the following way:

- Mgr. Alona Kononenko-Szoszkiewicz (100%): conceptualization, manuscript writing (preparation of the original article, reviewing and editing).

Authors:

Alona Kononenko-Szoszkiewicz

Mgr Alona Kononenko-Szoszkiewicz

PhD supervisor:

prof. dr hab. Katarzyna Dziubalska – Kołaczyk

Appendix 7. Research article 1: Dressler et al. (2019)

Dressler, Wolfgang U., Alona Kononenko, Sabine Sommer-Lolei, Katharina Korecky-Kröll, Paulina Zydorowicz & Laura Kamandulytė-Merfeldienė. 2019. "Morphological richness, transparency and the evolution of morphonotactic patterns", *Folia Linguistica Historica, Folia Linguistica* 40(1), 85-106.

DE GRUYTER MOUTON Folia Linguistica Historica 2019; 40(1): 85–106

Wolfgang U. Dressler*, Alona Kononenko,
Sabine Sommer-Lolei, Katharina Korecky-Kröll,
Paulina Zydorowicz and Laura Kamandulytė-Merfeldienė

Morphological richness, transparency and the evolution of morphonotactic patterns

https://doi.org/10.1515/flih-2019-0005
Received May 7, 2018; revised November 27, 2018; accepted December 21, 2018

Abstract: Morphonotactics determines phonological conditions on sound sequences produced by morphological operations both with morphemes and across boundaries. This paper examines the historical emergence and the development of morphonotactic consonant clusters in Germanic, Slavic, Baltic, Romance and other languages. It examines the role of the following morphological preference parameters: (i) morphotactic transparency/opacity, (ii) morphosemantic transparency/opacity, (iii) morphological richness. We identify several diachronic processes involved in cluster emergence, production and change: vowel loss, Indo-European ablaut (and comparable Arabic processes), affixation, compounding, metathesis, final and consonant epenthesis. Additionally, we discuss predictions derived from the Net Auditory Distance principle, psycholinguistic evidence and language acquisition. We show that the majority of morphonotactic clusters arise, phonologically, from vowel loss, and morphologically from concatenation.

Keywords: phonotactics, morphonotactics, consonant cluster emergence and development

1 Introduction

The sound patterns that a language admits in morphologically simple words often differ from those in complex words. In English, for example, no simple word ends

***Corresponding author: Wolfgang U. Dressler,** Department of Linguistics, University of Vienna, Vienna, Austria, E-mail: wolfgang.dressler@univie.ac.at
Alona Kononenko, Wydział Anglistyki, Uniwersytet im. Adama Mickiewicza, Poznań, Poland
Sabine Sommer-Lolei, Katharina Korecky-Kröll, Department of Linguistics, University of Vienna, Vienna, Austria
Paulina Zydorowicz, Wydział Anglistyki, Uniwersytet im. Adama Mickiewicza, Poznań, Poland
Laura Kamandulytė-Merfeldienė, Vytautas Magnus University, Kaunas, Lithuania

in /md/ or /vz/, while complex ones, such as *roamed*, or *lives* do. The patterns that occur in simple words reflect language-specific phonotactic constraints, and we call them 'phonotactic patterns' (PTs). Patterns produced in complex words, on the other hand, are referred to as 'morphonotactic patterns' (MPTs). MPTs often violate constraints that PTs obey. Clearly, some patterns may be PTs in some cases and MPTs in others, such as /nd/, which is PT in *wind*, and MPT in *sinned*.

In this paper, we focus on a subset of MPTs, namely consonant clusters. They often come about through affixation (e.g. *wife+s*, *ex+change*) or compounding (*black+board*), but also indirectly through morphologically motivated vowel deletion (Lat *spØr+ē+vi* 'despise 1SG.PERF' from *spern+o* 'despise 1SG.PRES'). MPT clusters reflect interactions between phonology and morphology. The traditional Neogrammarian position on such interactions was that phonological change (by sound law) was primary: it usually preceded and triggered morphological change (by analogy), which was therefore considered secondary (Hermann 1931). However, since Neogrammarian heydays many cases of primary morphological change have been found (Dressler 2002), and the ways in which phonology and morphology interact have turned out to be much more complex than envisaged (see Amdamczyk and Versloot (this issue)).

We consider evidence from various languages, and survey and classify the processes that produce and change clusters. Extending extant work in the area (Dressler et al. 2010; Baumann and Kaźmierski 2016), we compare historical developments to language acquisition and adduce evidence from language processing to arrive at explanatory accounts. In Section 2, we outline our theoretical background and our hypotheses, and introduce the concept of 'net auditory distance' (see also Dziubalska-Kołaczyk (this issue)), by which we assess the phonological preferability of consonant clusters. Section 3 surveys and classifies diachronic sources of clusters, Section 4 discusses phonological and psycholinguistic constraints on their emergence, and Section 5 provides a summary and an outlook.

2 Theoretical background and hypotheses

Our background is in Natural Linguistics (Dressler 1985, Dressler 1999), and specifically in Natural Phonology (Donegan and Stampe 1979; Donegan and Nathan 2015). Natural Linguistics deduces linguistic preferences from more general semiotic, cognitive or phonetic principles (Dressler 1999). To derive phonetically grounded preferences regarding consonant clusters, we use Dziubalska-Kołaczyk's Beats-and-Binding Model (2002, 2009, Forthcoming). There, the preferability of clusters is taken to reflect 'net auditory distances'

(NADs) between the involved consonants and their vocalic neighbours (Dziubalska-Kołaczyk 2002, and Forthcoming; Zydorowicz et al. 2016). This measure is both more comprehensive and more fine-grained than measures derived exclusively from the Sonority Sequencing principle (Jespersen 1904; Ohala and Kawasaki-Fukumori 1997; Ohala 2010; Dziubalska-Kołaczyk 2001).

Typologically, language-specific phonologies can be less or more consonantal. Their position on that scale reflects the size of their consonant inventory and the variety and complexity of the clusters they admit (cf. Maddieson 2013a; Donohue et al. 2013). On phonological grounds, clusters are the less preferred the more complex they are. Since processes that result in MPT clusters can disregard phonological preferences to some extent, MPT clusters are more likely to be phonologically dispreferred than PT clusters.

As far as morphological preference parameters (Dressler and Kilani-Schoch 2017) are concerned, the following ones are relevant for the present discussion:

a. morphotactic transparency/opacity
b. morphosemantic transparency/opacity
c. morphological richness

Morphotactic transparency refers to the ease with which the compositionality of a word form can be inferred from its sound shape. Full morphosemantic transparency means fully compositional meaning. This is generally the case in inflection, whereas in word formation compositional meanings may become opaque when they lexicalize and their morphological patterns become unproductive. Morphological richness refers to the wealth of productive morphological patterns in a language.

In our paper, we discuss the roles that preference parameters play in the development of morphonotactic clusters. Specifically, we discuss following six hypotheses:

H1: Due to the preference for morphotactic transparency more MPT clusters will come about through concatenation than through opacifying processes such as vowel deletion (as in Lat *spØr+ē+vi*, see above).

H2: The complexity of the MPT clusters that arise in a language will reflect how much complexity that language admits in clusters of any type, i.e. also in PT clusters.

H3: The richer the morphology of a language is, the more MPT clusters will arise. This generalisation is limited by a typological variable (in the sense of Skalička 1979): inflecting-fusional languages allow more morphotactic opacity than agglutinating languages.

H4: Morphological richness will also increase the chance of morphotactically opaque MPT clusters to arise.

H5: MPT clusters may become PT clusters when a word loses an internal morpheme boundary. This may happen in morphosemantically opaque words that are not token frequent, unless the pattern on which they are formed is highly productive.

H6: When MPT clusters become PT clusters, the change is always lexical and diffuses in steps, if at all. An MPT pattern never becomes PT in all forms that display it. This is related to the thesis that phonological rules can become morphonological rules (in the sense of Dressler 1985, Dressler 2002) but not *vice versa*.

3 Diachronic sources of consonant clusters

3.1 Vowel loss

Most PT clusters come about through vowel loss. Some of them may have MPT homophones that come about in the same way and that involve concatenation (cf. H1).

3.1.1 German

Good examples are German word final triple clusters ending in /st/. They derive from schwa loss in unstressed final syllables (Werner 1978). Two PT examples are *Herbst* 'fall' < Middle High German *Herbest* (cf. harvest), and *Ernst* 'seriousness' < MHG *Ernest* (cf. earnest). MPT clusters of that type involve the 2SG ending of verbs, as in *schimpf+st* 'scold-2SG' < MHG *schimpf+est*.

Schwa loss is often phonologically restricted. Before German word final sonorants it occurs only in casual speech. There, however, it also results in PT and MPT clusters, as in PT *Kolb(e)n* 'club, piston', *Wes(e)n* 'essence',[1] and MPT *geb+(e)n* 'give-INF/1PL.PRES/3PL.PRES', *les+(e)n* 'read-INF/1PL.PRES/3PL.PRES'.

There are also morphological restrictions (Stopp 1974; Thoursie 1984). German schwa loss did not occur in the subjunctive present (*schimpf+e+st* 'scold-SUBJ.PRES-2SG'), or in the preterite (*schimpf+te+st* 'scold-PST-2SG').

None of these schwa-deletions reduce morphotactic transparency, but in colloquial forms such as *geb+n* 'give-INF/1PL.PRES/3PL.PRES' and *leg+n* 'lay-INF/1PL.PRES/3PL.PRES' the nasals assimilate to preceding plosives, yielding [geːm] or [leːŋ], and this does reduce transparency.

1 See Stopp (1974), Issatschenko (1974), and Thoursie (1984) on phonological restrictions on schwa deletion.

In Southern German schwa reduction is blocked after nasals, as in *Rahmen* 'frame' or *wohn+en* 'live-INF/1PL.PRES/3PL.PRES'. This prevents cluster emergence and opacification, which both occur in casual colloquial Northern German, where forms such as *wohn+en* are reduced to [vo:n:] or even [vo:n].

Crucially, these changes affect all forms with the relevant patterns and have no lexical exceptions. When MPT clusters lose a morphological boundary and become PT clusters, on the other hand, this affects only individual words, never whole patterns (H5 and H6). It occurs when a complex word becomes opaque, or loses its base. Examples are *Kunst* 'art', whose relation to *könn+en* 'be capable' (cf. can) is no longer identified; *Oberst* 'colonel' (cf. *ober(er)* 'upper') whose superlative ending is no longer recognized; *zu+letzt* 'last' which has lost its positive and comparative counterparts; or *Kraft* 'strength', *Kluft* 'cleft', *Schaft* 'shaft', which were derived with a *-t* suffix that has become opaque.

In many German dialects the prefixes *be-* and *ge-* also lost their schwa before obstruents, and *ge-* also lost it before sonorants. So we have Austrian [ksofn] *ge+soff+en* 'drunk.PST.PTCP', and [psofn] *be+soff+en* 'drunk.ADJ'. Again, individual derivations may lose their boundaries yielding PT clusters as in [gmoa] *Gemeinde* 'municipality'. Also, the particle and prefix *zu-* lost its vowel before sonorants, as in [tsruk] *zurück* 'back(wards)', [tsletst] *zu+letzt* 'at last'. In [tsniaxtal] (< **zu-nicht-erl* [to-naught-DIM] 'unimportant person') the cluster has become PT through lexicalisation.

3.1.2 English

The diachrony of MPT clusters in English is similar to that of German (Baumann et al. 2015, Baumann et al. 2016; Baumann and Kaźmierski 2016), e.g. the PT cluster in OE *hand* (< *honed*) resulted from schwa loss, as did the MPT clusters in *sinn+ed* and the plural *sin+s*. Since English inflection is poorer than German inflection it created fewer MPT clusters. English has hardly any inflectional suffixes, and there is much homophony among the few that it has. For more on English phonotactics see the contributions by Baumann et al. (this issue); Honeybone, Minkova & Lefkowitz (this issue), and Schlüter (this issue).

3.1.3 Slavic languages

Also in Slavic languages, most PT and MPT clusters result from vowel loss. Their main source was the deletion (Walczak 1999: 45–46) of the two Proto-Slavic ultrashort high vowels *jer* (front) and *jor* (back) in unstressed position in the eleventh century. This deletion created novel PT clusters with no internal boundary, as in Proto-Slavic **pъtakъ* > Polish *ptak* 'bird'.

The deletion also created stem alternations that resulted in some opaque MPT clusters. Polish examples of root vowel deletion in oblique inflectional cases (here exemplified with the GEN.SG) are, *mech* – *mch+u* 'moss', *len* – *ln+u* 'linen', *wieś* – *ws+i* 'village', *wesz* – *wsz+y* 'louse', *lew* – *lw+a* 'lion', *bez* – *bz+u* 'elderberry', *kiep* – *kp+a* 'fool', *pień* – *pni+a* 'trunk', *szew* – *szw+u* 'stitch', *kieł* – *kł+a* 'fang' (*szw* and *kł* have PT homophones).

Vowel deletion occurred also in adjectives derived from nominal bases, as in *mch+owy* 'mossy', *lni+any* 'linen (ADJ)', *wsz+awy* 'lousy', *lw+i* 'leonine'.

Word-initial /mx/, /ln/, and /lv/ are always MPT. /ws/ and /kw/ are today also PT because some of the morphotactic operations that gave rise to them were lost.

Russian also has clusters resulting from vowel deletion, although the root vowel was retained or restored in many derivatives and compounds. The reason why vowel loss has been preserved in inflections is presumably that it is reflected in all oblique case forms, which may have created a "gang effect". See the examples in Table 1.

Table 1: Root with vowel deletion and without vowel deletion.

Root	With vowel deletion	Without vowel deletion
mox 'moss'	*mx+a* (GEN); *mš+istyj* (ADJ), *mš+al/vyj* (ADJ) 'covered with moss'	*mox+obraznyje* 'bryophytes'
lev 'lion'	*ľv+a* (GEN); *ľv+inyj* (ADJ), *ľv+-ica*, 'lioness', *ľv + enok* 'young lion'	–
led 'ice'	*ld+a* (GEN), *l'd+is* t- yj (ADJ), *ľd+ina*, 'ice-floe', *l'd+o+* generator 'icemaker'	*led+nik* 'glacier'; *led+o+bur* 'icebreaker', *led+o+xod* 'ice-drift', *led+o+*generator 'ice-maker'[a]
lob 'forehead'	*lb+a* (GEN)	*lob+ovoj* (ADJ) 'frontal', *lob+o +trjas*, 'idle', lit. 'head-shaking'
rož 'rye'	*rž+I* (GEN), *rž + ica* (DIM), *rž +išče* 'field from which the rye has been removed'	–
rot 'mouth'	*rt+a* (GEN)	*rot+ovoj* (ADJ) 'oral', *rot+ik* (DIM), *rot+o+zej* 'gaper'
rov 'ditch'	*rv+a* (GEN)	–
son 'dream'	*sn+a* (GEN)	*son+ny* (ADJ) 'sleepy'
šov 'stitch'	*š+va* (GEN)	–

[a]Not accepted as a standard variant.

In Standard Ukrainian, vowel deletion was sporadic in word initial syllables. Instead, the vowel that shows up as /o/ in oblique cases changed to /i/ in monosyllabic nominatives (Pugh and Press 2005), giving rise to alternations such as those between nominatives and genitives in *riv* – *rov+u* 'ditch', *vil* – *vol+a* 'ox', *bik* – *bok+u* 'side', mist – *most+u* 'bridge', etc. Thus, MPT clusters arose only in a few token-frequent words, such as *pes* – *ps+u* 'dog', *den'* – *dnj+a* 'day', *pen'* – *pnj+a* 'stump', *šov* – *šv+a* 'stitch'. In others, such as *lev* – *lev+a* 'lion' analogical levelling occurred.

In Slovak, the effects of root vowel deletion have been reversed by analogical levelling, e.g. in *mach* – *mach+u* (GEN) – *mach+ovy* (ADJ) 'moss', *ľan* – *ľan+u* (GEN) – ľan+ovy (ADJ) 'linen', *lev* – *lev+i* (GEN) – *lev+i* (ADJ) 'lion'. This re-established morphotactic transparency. It has been preserved in *pes* – *ps+y* (PL) 'dog' – *ps+í* (ADJ) – *ps+in+ec* 'dog den' – *ps+o+vod* 'dog guide'. This, we think, supports H4 that morphological richness makes morphotactically opaque MPT clusters more likely.[2] – However, most examples of root vowel deletion have been preserved in final unstressed syllables of disyllabic roots, such as *mozog* – *mozg+u* (GEN) 'brain', *laket'* – *lakt'+a* (GEN) 'elbow', *otec* – *otc+a* (GEN) 'father', *list+ok* – *list+k+a/u* (GEN/DIM) 'leaf'. Since Slovak has no word-final clusters *-zg, -tk, -tc, -stk*, morphonotactic root vowel deletion might have been reanalysed as phonotactic vowel insertion (cf. Dressler et al. 2015; Hliničanová et al. 2017). In any case, the stem alternations create some morphotactic opacity.

3.1.4 Baltic languages

In Lithuanian, a conservative Indo-European language, the only word-final MPT clusters that are due to vowel loss are in imperatives. The imperative suffix *-k* goes back to the particle *-ki*, and lost its final vowel only during the last centuries (Stang 1966: 219; Kazlauskas 1968: 382). The loss was sporadic and affected only *-ki*. It occurred when the particle grammaticalised into a suffix, and produced many clusters that are exclusively MPT, such as *dirb+k* 'work!', *temp+k* 'bend!', *megz+k* 'knit!', *skris+k* 'fly!', *lauž+k* 'break!', *blokš+k* 'give a blow!', *im+k* 'take!', *kel+k* 'lift!', *ar+k* 'plow!'

Word final /nk/, as in *aiškin+k* 'explain!', *sodin+k* 'plant!', on the other hand, also occurs in the mono-morphemic *link* 'towards'.[3]

2 Further, in the Ukrainian examples cited above, the words in which vowel deletion is preserved are highly token frequent.

3 Plus its derivations.

In Latvian, which is less conservative, vowel loss in unstressed final syllables has produced many final clusters, e.g. NOM.SG. *cilvek+s* 'man', *ak-men+s* 'stone', *rag+s* 'horn', *pil+s* 'castle', *av+s* 'sheep'.

3.1.5 Summary of cluster emergence through vowel loss

Clusters can clearly result from vowel deletion. The German examples in particular show how vowel deletion can create both MPT and PT clusters of the same structure (as in *Herbst* 'fall' and *stirb+st* 'die-2SG.PRES.IND'). There are, however, two different ways in which vowel deletion can produce MPT clusters. On the one hand, it can interact with concatenation, when a vowel in a -VC(C) suffix (such as German *-est* '2SG.PRES.IND') is deleted (= Scenario 1). In clusters arising that way, a morpheme boundary occurs between stem final consonant and the (first) consonant of the suffix. On the other hand, the effect of vowel deletion can be more indirect, as in the Slavic cases. There, vowels were deleted within roots, but only before certain suffixes or in certain compounds (= Scenario 2). Although the resulting clusters have no boundary inside, they are also morphologically conditioned. What is important is that the clusters arising in scenario 2 are predictably more diverse than the ones created in scenario 1, because suffixes belong to closed classes and are phonologically less diverse than lexical roots. That scenario 2 is attested in the morphologically rich Slavic languages but not in English or German, which are morphologically poorer, supports H3, which predicts a greater diversity of MPT clusters in morphologically richer languages. At the same time, the fact that there is no language in which clusters have emerged only in scenario 2, supports H1, which predicts that on the whole more MPT clusters will reflect concatenation than other morphological operations.

3.1.6 Ablaut

Phonological root vowel deletion must not be confused with morphological root vowel deletion in Indo-European zero grade ablaut, which also created MPT clusters as in (1):

(1) ...

a. /spr/ and /str/ in the Latin perfects *spØrē+vi*, *stØrā+vi*, and perfect participles *spØrē+tus*, *stØrā+tus* (< *sper+no* 'despise-1SG', *ster+no* 'strew.1SG')

b. /tm/ in the Ancient Greek perfect *té+tØmē-ka*, the passive aorist *e+tØmē +thēn*, and the verbal adjective
tØmē+tós (from *tém+no* ‘cut-1SG’), perfect *té+tmē-ka*
c. /tl/ in Ancient Greek in the verb = *tØlē+nai* = *talá+ssai* ‘suffer, endure’, which does not occur in the present, but has two aorists
d. /kt/ (a metathesis of /tk/) in the Ancient Greek zero grade reduplicated present *tí+kØt+o* (< *é+tek+on* ‘I gave birth to’

Quantitative ablaut has been speculated to derive, ultimately, from phonological vowel deletion in early Proto-Indo-European (Passler 1947; Mayrhofer-Passler 1952). Thus, ablaut patterns, which are purely morphological in attested languages and in reconstructible Proto-Indo-European, might ultimately also have a phonological origin.

The same applies to corresponding Semitic root patterns. While discontinuous *tri*- or *quadri*- consonantal roots are basic, they are unpronounceable, e.g. Ar. /ktb/ ‘write’. Thus, the lexical entry has been postulated to be either the 3SG.PRET *kataba* or its stem /katab/. This is warranted because the second vowel of the stem is unpredictable, as in the minimal triple *ḥasab-* ‘count’, *ḥasib-* ‘believe’, *ḥasub-* ‘be highly esteemed’ (for psycholinguistic evidence see Shimron 2002; Ravid et al. 2016). Thus, the MPT clusters in forms such as Ar. *a+ktub+u* ‘I write’, causative *a+ktab+a* ‘he dictated’, verbal noun *katb, kitb +a+t* ‘book’, *ma+ktab* ‘primary school’ are due to vowel loss. – Examples of diachronically attested vowel loss occur in modern Arabic varieties, notably in those of the Maghreb, such as in Tunisian Arabic, e.g. *ktāb* < *kitāb* ‘book’, *ma+ktb+a* < *ma+ktb+a+t* ‘library’ (Kilani-Schoch and Dressler 1985).

The synchronic analysis of “lacking vowels” in ablaut alternations may differ from diachronic vowel loss, but this holds also for many other cases of diachronic vowel loss, for which vowel epenthesis may be preferred in synchronic analyses.

3.2 Affixation

Clusters also arise when purely consonantal affixes combine with root-initial or root-final consonants. Although they are related to clusters that emerge when a suffix loses a vowel (as in German *schimpf+est* > *schimpf+st* ‘scold-2SG.PRES’ > *schimpf+st*), there is a difference in the order of sound change and morphological operation: the MPT clusters in this section come about through the affixation of morphemes that are already fully consonantal at the time of their use.

3.2.1 Italian

In Italian, such clusters arise word-initially through prefixation, as in *s+leale* 'illoyal', *s+qualificare* 'disqualify' (Iacobini 2004: 145–146, 159). s- goes back to Latin *ex-*, which lost its vowel and had the remaining /ks/ simplified to /s/ ([z] with voicing assimilation, see also Pustka's contribution to this issue). While the latter change was a regular sound law, the vowel loss was limited to the prefix. Some of the MPT clusters arising through *s*-prefixation have PT homophones, even complex ones like /skw/, which also occurs in words like *squadra* 'team'. Others, however, are exclusively MPT, such as [zr-, zdʒ̆-, zñ-], as in *sradicare* 'eradicate', *sgelare* 'defrost', *sgnaccare* 'to punish'. PT *sl-* occurs only in loan words (e.g. *slow food*), and the fact that it occurs at all is clearly due to the MPT model.

3.2.2 Slavic languages

In Slavic languages, consonantal prefixes were produced by the Proto-Slavic loss of the ultrashort high back vowel (Walczak 1999: 45–46, see also above). Ultimately, these prefixes may go back to Indo-European "preverbs" (Watkins 1963): preverbs were particles, which Proto-Indo-European had instead of verbal prefixes. Morphological and syntactic elements could be inserted between them and following verb stems. Slavic prefixes arose only later through univerbation, which occurred before the two Proto-Slavic prefixes lost their unstressed vowels. It represents a precondition for the emergence of the word initial MPT-clusters attested in modern Slavic languages.

In Polish, the prefixes *w-* ([v] or [f] by devoicing) and *s-*, and their combination *w-s-* ([fs]) have given rise to complex clusters, many of which are exclusively MPT, such as all word-initial *ws-* clusters and all quadruple clusters (e.g. *ws+tręt* [fstr-] 'disgust' and *w+strzelać* 'to shoot in').

Cognate Russian prefixes create word-initial MPT clusters beginning with [v] or [f]. They include twelve quadruple clusters: /vzbr/, /vzgl/, /vzgr/, /vzdr/, /fskl/, /vzdv/, /fskr/, /fspl/, /fspr/, /fstr/, /vsxl/, /vsxr/, with only few phonotactic homophones, such as /fstr/ in *fstreča* 'meeting'.

Also in Ukrainian, there exist word-initial quadruple clusters such as /vzdr/ in the dialectal perfective verb *v+z+driv* 'has seen', /vpxn/ in the imperative *v+pxny* 'shove sth. in!', /vstr/ in dialectal *v+striv* 'has met', /vškv/ in *v+škvaryty* 'to strike'. They are exclusively MPT.

In Slovak, the word-initial clusters /vb/, /vp/, and /vst/ as in *v+bit'* 'to hit on', *v+padnúť* 'to fall in', and *v+stat'* 'to get up' are exclusively MPT, but both in Slovak and Polish the majority of MPT clusters has PT homophones. This seems to support H2, predicting that the complexity of the MPT clusters arising in a language will reflect the complexity that this language admits in general.

3.2.3 Latin, Lithuanian, Latvian

In conservative Indo-European languages word-final clusters ending in /s/ occur almost exclusively in inflected forms, e.g. in the Latin NOM.SG in

(6) a. /n+s/: *lauda+n+s* 'praising', *fon+s* 'source', *den+s* 'tooth', *pon+s* 'bridge' (vs. PT /ns/ in *trans* 'beyond'), all with loss of the stem-final dental stop
b. /k+s/: *pax* 'peace', *dux* 'leader', *lex* 'law' (vs. rare PT /ks/ in *sex* '6' *mox* 'soon')
c. /p+s/: *inop+s* 'helpless', *pleb+s* (vs. isolated PT /ps/ in *abs* = *ab* 'away')

After sonorants, the resulting triples are exclusively MPT, as in *stirps*, *arx*, *falx*, *lanx*.

Note that the PT cluster in *trans* supports H6, which predicts that changes from MPT to PT clusters diffuse lexically and in steps: *trans* goes back to the present participle of extinct *trare* (preserved in *intrare* 'enter'), but has lost its boundary in lexicalisation. It is the only item in which a PT /ns/ cluster came about in that way.

Lithuanian MPT clusters (except the ones in Section 3.1.4) result from the attachment of consonantal affixes. One example is future formation, as in the 3.FUT forms *kep+s* 'bake', *dirb+s* ' work', *kel+s* 'lift', *gin+s* 'defend', or *megz+s* [meks] 'knit'.[4] Other clusters occur in irregular genitive singulars, where *-s* is added and which tend to be replaced, in colloquial Lithuanian, by productive genitives without clusters, as in the genitives *obel+s* > *obel+ies* 'apple tree', *moter+s* > *moter+ies* 'woman', *šun+s* > *šun+io* 'dog', *piemen+s* > *piemen+io* 'shepherd' (cf. Ambrazas 2006: 79–80).

3.2.4 The typological variable of morphotactic opacity

The morphologically richer Slavic languages have more MPT clusters than the morphologically poorer Germanic and Romance languages. This supports the first part of H3, which predicts more MPT clusters for morphologically richer languages.

4 *megz+s*, imperative *megz+k*, and infinitive *megz+ti* reflect metathesis of root-final /zg/ before consonants.

The second part of H3 predicts that more MPT clusters in inflecting-fusional languages than in agglutinative ones. It is supported by the small number of MPT clusters in morphologically rich agglutinating languages such as Finnish. In Finnish no MPT clusters result from affixation (*ex+puoliso* 'former husband') or compounding (*syys+myrsky* 'autumn storm'), but MPT clusters are often simplified through consonant deletion or assimilation, as in the partitive *last +ta* of *lapsi* 'child', or the participle *hakan+nut* of *hakata* 'to beat' (Klaus Laalo, pers. comm.). Estonian, which is less agglutinating than Finnish (Skalička 1979), has more vowel loss and also more clusters (Skalička 1979: 308).

Agglutinating Hungarian often avoids MPT clusters by vowel insertion, as in PL. *biciklis+ek* 'bicycles', ACC. *Szék+et* 'chair', LOC. *Pécs+ett* 'in Pécs', iterative *het+ente/nap+onta* 'every week/day'. Still, Hungarian has many PT clusters (although word-initially only in loans). It also has word-final and word-medial MPTs (Kenesei et al. 1998: 386–409). It is not clear to what extent it supports H3.

3.2.5 Summary of cluster emergence through affixation

Affixation can create diverse and complex types of MPT clusters. Among the languages we have looked at, the greatest diversity seems to be attested in Slavic languages, which are morphologically rich and allow diverse and complex PT clusters as well. This confirms our expectations. We have also shown, again, that changes of MPT clusters into PT clusters always affect only individual lexical items (H6).

3.3 Compounding (German and Lithuanian)

German is particularly rich in compounds, and compounding is a source of many German MPT clusters (Dressler et al. 2015; Dressler and Kononenko 2018). Many German compounds are formed with the interfix *-s-* as in *König+s+hof* 'royal court, lit. 'king+INTERFIX+court'. This obviously increases the complexity of the word internal clusters. In this respect, the Germanic family differs from other Indo-European languages, where interfixes are typically vocalic, highly productive, often obligatory, and prevent clusters rather than increasing their complexity. An example would be the interfix *-o-* in words such as *gas+o+meter*.

Also in Lithuanian, word-internal MPT clusters arise in compounding, when the stem-final thematic vowel is deleted and the bare root appears as first part of a compound. Of the 80 MPT clusters arising in Lithuanian compounding, 46 have PT homophones and 34 are exclusively MPT (Dressler et al. 2010). Some examples are:

(7) /bg/: *sklyp* + *gal -i -s* 'end of a plot' ← *sklyp-a-s* 'plot' + *gal-a-s* 'end'
/tp/: *rud[a]* + *plauk-i-s* 'brown-haired' ←*rud-a-s* 'brown' + *plauk-ai* 'hair' (Pl.)
/tk/: *led(a)* + *kaln-i-s* 'iceberg' ←*led-a-s* 'ice' + *kaln-a-s* 'mountain'
/gb/: *pilk* + *balt-i-s* 'grey-white' ←*pilk-a-s* 'grey' + *balt-a-s* 'white'
/čt/: *didž* + *turt-i-s* 'wealthy man' ←*did-i-s* 'big' + *turt-a-s* 'wealth'
/džd/: *treč* + *dal-i-s* 'one third' ←*treči-a* 'third' + *dal-i-s* 'part'
/džg/: *plač*/t + *gal-y-s* 'oar blade' ←*plat-u-s* 'broad' + *gal-a-s* 'end'
/fs/: *gyv* + *sidabr-i-s* 'mercury' ←*gyv-a-s* 'lively' + *sidabr-a-s* 'silver'
/fš/: *diev* +š *auk+i+s* 'praying person' ←*diev-a-s* 'god' + *šauk-ti* 'to call'
/šs/: *kryž* + *snap-i-s* 'crossbill' ←*kryži-u-s* 'cross' + *snap-a-s* 'beacon'

3.4 Other sources

3.4.1 Metathesis

Consonant clusters can also result from metathesis, such as Slavic liquid metathesis (Boryś 2005), as in Polish *mleko* (< **melko*) 'milk', or *groch* (< **gorchъ*) > 'pea'.

3.4.2 Final consonant epenthesis

An exceptional case of cluster emergence occurred in the German words *Axt* 'axe' and *Obst* 'fruit', and *Palast* 'palace'. The clusters result from an 'unetymological' addition of word final /t/, which may have prevented the final *-s* from being mistaken for the frequent genitive suffix *-s*. Also, /st/ (as in *Mast* 'mast', or *Forst* 'forest') was a much more common noun ending than /s/. Furthermore, *-st* was a nominalizing suffix (as in *Dien+st* 'service' from *dienen* 'serve'). Thus, the addition of /t/ might have made the nouns more easily recognizable as nouns. However, this does not fully explain similar examples such as *Sekt* 'sparkling wine' (< Fr *vin sec*), *Werft* 'shipyard' (cf. E *wharf*), and dialectal Austrian German *Senft* (< Senf 'mustard'). All one can observe is that the resulting word-final clusters had PT and MPT models, and that German generally allows more complex clusters at the ends of words than at their beginnings (Dressler and Kononenko 2018). This can be understood as partial support of hypothesis H2, which predicts that the complexity of the MPT clusters a language allows should correlate with the cluster complexity it allows generally i.e. also in PT clusters.

3.4.3 Internal consonant epenthesis

Internal consonant epenthesis is a common transition phenomenon. It occurs in Austrian German dialects in diminutives with the suffix /l/, as in *Mandl* from *Mann* 'man', or *Hendl* from *Henne* 'hen'. /ndl/ is an MPT cluster, but when the base *Henne* was lost, *Hendl* ceased to be a diminutive and the cluster became PT. Dental epenthesis has also created the initial clusters in G *Strom* and E 'stream' (< **sroumo-*) and in Pol *strumień* 'creek'. In North Mazovian dialects of Polish epenthesis produced /hɛndrɨk/ = Standard /hɛnrɨk/ *Henryk* 'Henry', /tšɛrɛmpka/ = Standard /tšɛremxa/ *czeremcha* 'bird cherry', /rustcɛ/ = Standard /ruscɛ/ *Ruskie* 'Russians (derogatory)' (Czaplicki 2010).

Often vowel loss and consonant epenthesis combine, as in Proto-Slavic **bъčela* > Pol *pszczoła* 'bee', Lat *ponere* 'to put' > Fr *pondre* 'lay eggs', Lat *cinerem* 'ash (ACC)' > Fr *cendre*, Lat *hominem* 'man (ACC)' > Sp *Hombre*.

Labial epenthesis between root-final /m/ and a dental suffix has taken place in G *Brunft* 'rutting season', *(An)kunft* 'arrival', *Vernunft* 'reason', *Zunft* 'guild'. When the words became morphosemantically opaque, the MPT clusters became PT.

4 Explaining the diachronic emergence and the historical stability of MPT clusters

The survey in Section 3 has been primarily descriptive, although we have related our findings to hypotheses when appropriate. In this section, we discuss predictions about specific properties to be expected from MPT clusters and about differences between MPT and PT clusters. Some predictions are derived from phonetically grounded preferences, others from psycholinguistic evidence and language acquisition.

4.1 The impact of phonological preferences (Net Auditory Distance)

Since clusters are generally not preferred, one hypothesis is that MPT clusters should be phonologically even less preferred than PT clusters. This is because (a) MPT clusters can signal boundaries, particularly when they have no PT homophones, (b) they often include consonantal morphemes, which are motivated by their morphosemantic functions. This may outweigh the articulatory and perceptual difficulties that MPT clusters create. Thus, MPT clusters 'can afford' to be less preferred than PT clusters, and therefore will be.

One way to assess the relative preferability of clusters is in terms of their Net Auditory Distance profile, described in Dziubalska-Kołaczyk's contribution to this issue. Applying this measure to the MPT clusters in the languages discussed above, it turns out that the majority of MPT clusters are indeed dispreferred in Polish and Russian, but not in German and Austrian German. It is therefore premature to decide on the hypothesis. Russian and Polish may be more suitable as test cases, because their cluster inventory is greater (e.g. more than 100 word initial triple clusters in Polish, only 8 in German), and may yield more significant results. At the same time, the preferability of MPTs clusters does not seem to affect their chance of being ousted in analogical levelling.

Clearly, the complexity that a language allows in clusters is related to the complexity it allows in syllables (Maddieson 2013b; Duanmu 2008). Since clusters may span syllable boundaries, word-internal ones will be complex more often than peripheral ones. Syllable structure can also explain, at least partially, why some languages (e.g. Slavic languages) allow more complexity initially than finally and why the opposite is true of others (e.g. German): the Slavic distribution reflects a strong preference for open syllables attested already in earliest stages.

4.2 Language acquisition and psycholinguistic factors

The status of MPT clusters in language acquisition is relevant for diachrony, as (a) at least some diachronic changes may come about through 'imperfect', non-target-like acquisition (Dressler 1997), and (b) early acquisition supports historical stability. Moreover, (c), morphological richness stimulates and facilitates the acquisition of morphology (Xanthos et al. 2011). Here we briefly report some of our studies on the acquisition of MPT versus PT clusters and discuss what they might mean for their histories.

Freiberger (2014) investigated a longitudinal corpus of spontaneous mother-child interaction of three monolingual toddlers aged 1;7 to 3;0, acquiring German. She found a significant effect of position, but the ages at which MPT and PT clusters were acquired did not differ significantly. The same result was obtained by Korecky-Kröll et al. (2015, 2016): longitudinal spontaneous parent-child interaction data of slightly older children (3;0–5;0) showed that socio-economic status was relevant, but not the difference between MPT and PT clusters. In a comparative study, on the other hand (Zydorowicz et al. 2015), we found that Polish and Lithuanian children acquired (i.e. produced) MPT clusters significantly earlier than PT homophones, while English and German children did not do so. This once again suggests that the morphology of a language needs to be sufficiently rich, and its cluster inventory sufficiently large, for differences between MPT and PT patterns to become significant.

As far as cluster emergence through vowel loss is concerned, reduction processes can in principle occur in all age brackets and are usually insensitive to morphological boundaries. Still, acquisition may explain why diachronic vowel loss affected the verbal 2SG only in the present indicative (*schimpf+st* 'scold-2SG.PRES.IND'), but not in the subjunctive (*schimpf+est* 'scold-2SG.PRES.SUBJ') or the verbal past (*schimpf+t+est* 'scold-PST.2SG'). Recency effects in early learning diminish over time. Thus, young children abbreviate Ger. *Schokolade* 'chocolate' as *Ladi* ['la:dɪ] (recency effect), but adults as *Schoko* [ˈʃo:ko] (primacy effect). Since the present indicative is acquired earlier than the subjunctive and preterite, this may cause early entrenchment of reduced, vowel-less /st/. At the same time, the fact that already young children try to maintain morphotactic contrasts between morphosemantically contrasting categories may explain why they retained the vowel in Early Modern High German preterites and subjunctives.

Analogy in language acquisition may be responsible for blocking the simplification of MPT clusters in forms like *haben* 'have-INF' [ˈha:bm̩] to [ha:m̩] . Children are likely to prefer the disyllabic pattern because it is more transparent and makes complexity easier to identify.

Experiments about processing in adolescents and adults have also yielded ambivalent results. In a letter decision task (Korecky-Kröll et al. 2014), the morpheme boundary in MPT clusters turned out to be helpful. In a fragment monitoring task (Celata et al. 2015), on the other hand, both adults and adolescents were significantly faster in detecting sequences containing PT than MPT clusters, and with respect to accuracy there were no significant differences. In a split-cluster task (Celata et al. 2015), finally, adolescents, but not adults, split significantly more MPT than PT clusters in an accurate way.

Studies addressing the lexical level show that complex words with MPT clusters are processed more slowly than simple ones with PT clusters (Freiberger et al. 2015). This supports previous findings that have identified higher processing costs in inflected word forms as opposed to monomorphemic words (e.g. Laine et al. 1999). It says little about the difference between PT and MPT clusters, however. Therefore, we investigated a domain where German morphology is rich, namely compounding (Sommer-Lolei et al. 2018). We conducted a lexical decision task that contrasted German compounds with monomorphemic nouns, both with and without clusters, e.g. *Haus+tier* 'domestic animal' vs. *Tee+tasse* 'teacup' vs. *Kastanie* 'chestnut' vs. *Rakete* 'rocket'. We found a significantly higher accuracy for compounds with clusters at the boundary than for all other types of stimuli. As far as reaction time was concerned, the trend was the same, but statistically not significant. Thus, our results suggest that significant differences between MPT and PT clusters show up only in domains where the inventory of clusters that the morphology of a language produces is sufficiently rich. The results from processing experiments in these domains, however, do predict a greater diversity of historically stable MPT clusters in languages that are rich in morphology (Sommer-Lolei et al. 2018; Zydorowicz et al. 2015), and in clusters.[5]

5 Summary and outlook

Most MPT clusters arise in one of two ways: phonologically, they arise through vowel loss, and morphologically, they arise through concatenation (H1). These two possibilities intersect orthogonally, especially in affixation of suffixes that have come to be consonantal or that provoke vowel loss. In contrast, the complexity of MPT clusters and their position in the word is primarily due to the phonological complexity of syllable structure (H2) and only secondarily to the richness of morphology (H3, H4). An intervening variable is the degree of morphotactic opacity that the language type allows (H3). MPT clusters may become PT clusters only via lexical change (H5, H6). This reflects the semiotic priority of the lexicon over phonology (Dressler 1985, Dressler 2002; Dressler and Kilani-Schoch 2017).

A question that our observations raise is whether MPT clusters differ from PT clusters as far as their emergence, their historical stability, or their loss are concerned: do their histories reflect, at least partly, their specific status in an area where phonology and morphology overlap and interact, i.e. do they have histories in their own right? Or do their histories represent mere epiphenomena of developments that happen, irrespectively, to sounds, on the one hand, and to morphemes on the other?

Our typological comparisons and our evidence from language acquisition and processing suggest that differences affecting the histories of MPT clusters and PT clusters are slight and difficult to detect. In languages (or subdomains such as German compounding) that are both morphologically rich and rich in consonants, however, it seems that MPT clusters may indeed acquire a status that distinguishes them from PT clusters. In such languages (or domains) evidence from typology (which both emerges from and shapes the historical evolution of languages), from acquisition (Freiberger 2014; Kamandulytė

5 At least Wagner et al. (2012) have found that an existing cluster can be better perceived in a language rich in clusters [Polish] than in a language [English] poorer in clusters.

2006; Kamandulytė-Merfeldienė 2015; Korecky-Kröll et al. 2016; Sommer-Lolei et al. 2018; Zydorowicz 2007, Zydorowicz 2010; Zydorowicz et al. 2015), and from processing converge. It suggests that a sufficiently great diversity of clusters and morphological operations seem to be required for speakers to become sensitive to systematic distinctions between MPT clusters and PT clusters, to make respective generalizations and abstractions, exploit them in learning, processing and use, and to transmit them stably across generations, thereby establishing MPT clusters in their languages.

Although our conclusions need to remain tentative, they suggest in which languages and domains further research on morphonotactic patterns and their histories promises to be particularly productive.

Acknowledgements: We thank Chiara Celata (Pisa), Klaus Laalo (Tampere), Paula Orzechowska (Poznań), Martin Peters (Vienna) and Anke Sennema (Vienna), two anonymous reviewers and the editors for their help and valuable suggestions.

References

Ambrazas, Vytautas (ed.). 2006. *Dabartinės lietuvių kalbos gramatika* [A grammar of modern Lithuanian]. Vilnius: Mokslo ir enciklopedijų leidybos institutas.

Baumann, Andreas & Kamil Kaźmierski. 2016. A dynamical-systems approach to the evolution of morphonotactic and lexical consonant clusters in English and Polish. *Yearbook of the Poznan Linguistic Meeting* 2. 115–139.

Baumann, Andreas, Christina Prömer & Nikolaus Ritt. 2015. Identifying therapeutic changes by simulating virtual language stages: A method and its application to Middle English coda phonotactics after schwa deletion. *VIEWS* 24. 1–31.

Baumann, Andreas, Christina Prömer & Nikolaus Ritt. 2016. Diachronic dynamics of Middle English phonotactics provide evidence for analogy effects among lexical and morphonotactic consonant clusters. *Papers in Historical Phonology* 1. 50–75.

Boryś, Wiesław. 2005. *Słownik etymologiczny języka polskiego* [Polish etymological dictionary]. Kraków: Wydawnictwo Litterackie.

Celata, Chiara, Katharina Korecky-Kröll, Irene Ricci & Wolfgang U. Dressler. 2015. Phonotactic processing and morpheme boundaries: Word-final /Cst/ clusters in German. *Italian Journal of Linguistics* 27. 85–110.

Czaplicki, Bartłomiej. 2010. Emergent stops in English and in Polish: Against syllable-based accounts. *Poznan Studies in Contemporary Linguistics* 46. 177–191.

Donegan, Patricia J. & Geoffrey S. Nathan. 2015. Natural phonology and sound change. In Patrick Honeybone & Joseph Salmons (eds.), *Oxford handbook of historical phonology*, 431–449. Oxford: Oxford University Press.

Donegan, Patricia J. & David Stampe. 1979. The study of natural phonology. In D. A. Dinnsen (ed.), *Current approaches to phonological theory*, 126–173. Bloomington: Indiana University Press.

Donohue, Mark, Rebecca Hetherington, James Mc Elvenny & Virginia Dawson. 2013. *World phonotactics data-base*. Canberra: Department of Linguistics, The Australian National University.
Dressler, Wolfgang & Alona Kononenko. 2018. German phonotactic vs. morphonotactic obstruent clusters: A corpuslinguistic analysis. In Wolfgang U. Dressler, Sabine Sommer-Lolei & Basilio Calderone (eds.), *Experimental and acquisitional approaches to morphonotactics*. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.
Dressler, Wolfgang U. 1985. *Morphonology. The dynamics of derivation*. Ann Arbor: Karoma.
Dressler, Wolfgang U. 1997. "Scenario" as a concept for the functional explanation of language change. In Jadranka Gvozdanović (ed.), *Language change and functional explanations*, 109–142. Berlin: Mouton de Gruyter.
Dressler, Wolfgang U. 1999. On a semiotic theory of preferences in language. In Michael Shapiro (ed.), *The Peirce seminar papers*, vol. *4*. 389–415. New York: Bergham Books.
Dressler, Wolfgang U. 2002. Naturalness and morphological change. In Brian D. Joseph & Richard D. Janda (eds.), *The handbook of historical linguistics*, 461–471. Oxford: Blackwell.
Dressler, Wolfgang U., Katarzyna Dziubalska-Kołaczyk & Lina Pestal. 2010. Change and variation in morphonotactics. *Folia Linguistica Historica* 31(1). 51–67.
Dressler, Wolfgang U., Miroslava Hliničanová, Matej Ďurčo, Karlheinz Mörth & Katharina Korecky-Kröll. 2015. Phonotactic vs. morphonotactic obstruent clusters in Slovak and German. *Italian Journal of Linguistics* 27. 45–60.
Dressler, Wolfgang U. & Marianne Kilani-Schoch. 2017. Natural morphology. In Andrew Hippisley & Gregory Stump (eds.), *The Cambridge handbook of morphology*, 356–389. Cambridge: Cambridge University Press.
Duanmu, San. 2008. *Syllable structure: The limits of variation*. Oxford: Oxford University Press.
Dziubalska-Kołaczyk, Katarzyna. 2001. Phonotactic constraints are preferences. In Katarzyna Dziubalska-Kołaczyk (ed.), *Constraints and preferences*, 69–100. Berlin: Mouton de Gruyter.
Dziubalska-Kołaczyk, Katarzyna. 2002. *Beats-and-binding phonology*. Frankfurt am Main: Peter Lang.
Dziubalska-Kołaczyk, Katarzyna. 2009. NP extension: B&B phonotactics. *Poznan Studies in Contemporary Linguistics* 45(1). 55–71.
Freiberger, Eva Maria. 2014. Der Erwerb von Konsonantenclustern im Deutschen: Eine Untersuchung longitudinaler Spontansprachkorpora [The acquisition of consonant clusters in German; an investigation of longitudinal spontaneous corpora]. *Slowakische Zeitschrift Für Germanistik* 6. 5–23.
Freiberger, Eva Maria, Katharina Korecky-Kröll, Basilio Calderone & Wolfgang U. Dressler. 2015. Morpheme boundaries inside consonant clusters delay visual word recognition: Evidence from a lexical decision and a progressive demasking experiment. Paper presented at the 3rd International Workshop on Phonotactics and Phonotactic Modeling (PPM 2015), Vienna, 26 November.
Hermann, Eduard. 1931. *Lautgesetz und Analogie*. Berlin: Weidmann.
Hliničanová, Miroslava, Matej Ďurčo, Karlheinz Mörth & Wolfgang U. Dressler. 2017. Phonotaktische versus morphonotaktische Konsonantengruppen im Slowakischen und Deutschen: Eine kontrastive korpuslinguistische Untersuchung. In Claudia Resch & Wolfgang U. Dressler (eds.), *Digitale Methoden der Korpusforschung in Österreich (Veröffentlichungen zur Linguistik und Kommunikationsforschung Nr. 30)*, 159–178. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.

Iacobini, Claudio. 2004. Prefissazione [Prefixation]. In Maria Grossmann & Franz Rainer (eds.), *La formazione delle parole in Italiano*, 69–163. Tübingen: Niemeyer.

Issatschenko, Alexander. 1974. Das "Schwa mobile" und "Schwa constans" im Deutschen [„Mobile schwa" and „constant schwa" in German]. In Ulrich Engel & Paul Grebe (eds.), *Sprachsystem und Sprachgebrauch, Festschrift für Hugo Moser*, vol. *1*, 142–171. Düsseldorf: Pädagogischer Verlag Schwann.

Jespersen, Otto. 1904. *Lehrbuch der Phonetik*. Leipzig: Teubner.

Kamandulyte, Laura. 2006. The acquisition of morphonotactics in Lithuanian. *Wiener Linguistische Gazette* 73. 88–96.

Kamandulytė-Merfeldienė, Laura. 2015. Morphonotactics in L1 acquisition of Lithuanian: TD vs SLI. *Eesti Rakenduslingvistika Ühingu Aastaraamat [Estonian Papers in Applied Linguistics]* 11. 95–109.

Kazlauskas, Jonas. 1968. *Lietuvių kalbos istorinė gramatika* [Historical grammar of the Lithuanian language]. Vilnius: Mintis.

Kenesei, István, Robert M. Vago & Anna Fenyvesi. 1998. *Hungarian*. London: Routledge.

Kilani-Schoch, Marianne & Wolfgang U. Dressler. 1985. Natural morphology and classical vs. Tunisian Arabic. *Studia Gramatyczne* 7. 27–47.

Korecky-Kröll, Katharina & Wolfgang U. Dressler. 2015. (Mor)phonotactics in high vs. low SES children. In Katarzyna Dziubalska-Kołaczyk & Jarosław Weckwerth (eds.), *Rajendra Singh: In memoriam. Papers from a special commemorative session at the 44th Poznań Linguistic Meeting*, 25–40. Poznań: Adam Mickiewicz University Press.

Korecky-Kröll, Katharina, Wolfgang U. Dressler, Eva Maria Freiberger, Eva Reinisch, Karlheinz Mörth & Gary Libben. 2014. Morphonotactic and phonotactic processing in German-speaking adults. *Language Sciences* 46. 48–58.

Korecky-Kröll, Katharina, Eva M. Freiberger, Wolfgang U. Dressler & Basilio Calderone. 2016. Acquiring and processing German words of different (mor)phonotactic structure: Morpheme boundaries inside consonant clusters delay processing, but not acquisition. Poster presented at the 10th International Conference on the Mental Lexicon. Ottawa, 19 October.

Laine, Matti, Seppo Vainio & Hyönä Jukka. 1999. Lexical access routes to nouns in a morphologically rich language. *Journal of Memory and Language* 40. 109–135.

Maddieson, Ian. 2013a. Consonant inventories. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/1 accessed on 23/4/18.

Maddieson, Ian. 2013b. Syllable structure. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/1 accessed on 1/11/18.

Mayrhofer-Passler, Edeltraut. 1952. Der Quantitätsablaut in den indogermanischen Sprachen [Quantitative apophony in the Indo-European languages]. In Wilhelm Brandenstein (ed.), *Studien zur indogermanischen Grundsprache*, 15–22. Wien: Gerold.

Ohala, John. 2010. The emergent syllable. In Barbara L. David & Krisztina Zajdó (eds.), *The syllables in speech production*, 179–186. New York: Lawrence Erlbaum Associates.

Ohala, John & Haruko Kawasaki-Fukumori. 1997. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In Stig Eliasson & Ernst Håkon Jahr (eds.), *Language and its ecology: Essays in memory of Einar Haugen. Trends in linguistics. Studies and monographs*, vol. *100*. 343–365. Berlin: Mouton de Gruyter.

Passler, Edeltraut. 1947. Der indogermanische Ablaut als funktionelles Element [The Indo-European apophony as functional element]. *Revue Des Études Indo-Européennes* 4. 365–404.
Pugh, Stephan & Ian Press. 2005. *Ukrainian: A comprehensive grammar*. London: Routledge.
Ravid, Dorit, Orit Ashkenazi, Ronit Levie, G. Ben-Zadok, T. Grunwald & Steven Gillis. 2016. Foundations of the early root category: Analyses of linguistic input to Hebrew-speaking children. In Ruth A. Berman (ed.), *Acquisition and development of Hebrew: From infancy to adolescence*, 95–134. Amsterdam: Benjamins.
Shimron, Joseph (ed.). 2002. *Language processing and acquisition in languages of Semitic, root-based, morphology*. Amsterdam: Benjamins.
Skalička, Vladimír. 1979. *Typologische Studien*. Braunschweig: Viewegh.
Sommer-Lolei, Sabine, Katharina Korecky-Kröll & Wolfgang U. Dressler. 2018. Acquisition and processing of (mor)phonotactic patterns of German consonant clusters. In Wolfgang U. Dressler, Sabine Sommer-Lolei & Basilio Calderone (eds.), *Experimental and acquisitional approaches to Morphonotactics*. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.
Stang, Christian. 1966. *Vergleichende Grammatik der Baltischen Sprachen [*Comparative grammar of the Baltic languages]. Oslo: Universitetsforlaget.
Stopp, Hugo. 1974. Veränderungen im System der Substantivflexion vom Althochdeutschen bis zum Neuhochdeutschen [Changes in the system of noun inflection from old high German to new high German]. In Ulrich Engel & Paul Grebe (eds.), *Sprachsystem und Sprachgebrauch, Festschrift für Hugo Moser*, vol. 2, 324–344. Düsseldorf: Pädagogischer Verlag Schwann.
Thoursie, Stig A. O. 1984. *Die Verbalflexion eines südbairischen Autographs aus dem Jahr 1464, Ein Beitrag zur frühneuhochdeutschen Morphologie* [Verbal inflection of a Southern Bavarian manuscript of 1464. A contribution to early New High German morphology]. Göteborg: Acta Universitatis Gothoburgensis.
Wagner, Monica, Valerie L. Shafer, Brett Martin & Mitchell Steinschneider. 2012. The phonotactic influence on the perception of a consonant cluster /pt/ by native English and native Polish listeners: A behavioural and event related potential (ERP) study. *Brain and Language* 123. 30–41.
Walczak, Bogdan. 1999. *Zarys dziejów języka polskiego*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.
Watkins, Calvert. 1963. Preliminaries to a historical and comparative analysis of the syntax of the old Irish verb. *Celtica* 6. 1–49.
Werner, Otmar. 1978. Schwa-Schwund und Phonotaktik im Deutschen. In H.D. Pohl (ed.), *Opuscula Slavica et linguistica: Festschrift für Alexander Issatschenko*, 471–486. Klagenfurt: Heyn.
Xanthos, Aris, Sabine Laaha, Steven Gillis, Ursula Stephany, Ayhan Aksu-Koç, Anastasia Christofidou, Natalia Gagarina, Gordana Hrzica, F. Nihan Ketrez, Marianne Kilani-Schoch, Katharina Korecky-Kröll, Melita Kovačević, Klaus Laalo, Marijan Palmović, Barbara Pfeiler, Maria D. Voeikova & Wolfgang U. Dressler. 2011. On the role of morphological richness in the early development of noun and verb inflection. *First Language* 31. 461–479.
Zydorowicz, Paulina. 2007. Polish morphonotactics in first language acquisition. *Wiener Linguistische Gazette* 74. 24–44.
Zydorowicz, Paulina. 2010. Consonant clusters across morpheme boundaries: Polish morphonotactic inventory and its acquisition. *Papers and Studies in Contemporary Linguistics* 46. 565–588.

Zydorowicz, Paulina, Laura Kamandulyte, Eva Freiberger, Katharina Korecky-Kröll & Wolfgang U. Dressler. 2015. Morphological richness facilitates first language acquisition of morphonotactic consonant clusters: Polish and Lithuanian vs. Austrian German and English. *Paper presented at the 3rd International Workshop on Phonotactics and Phonotactic Modeling (PPM 2015)*. Vienna, 26th November.

Zydorowicz, Paulina, Paula Orzechowska, Michał Jankowski, Katarzyna Dziubalska-Kołaczyk, Piotr Wierzchoń & Dawid Pietrala. 2016. *Phonotactics and morphonotactics of Polish and English: Description, tools and applications*. Poznań: Wydawnictwo Naukowe UAM.

Appendix 8. Research article 2: Dressler & Kononenko-Szoszkiewicz, 2019)

Dressler, Wolfgang U. and Alona Kononenko-Szoszkiewicz. 2019. "Main Differences Between German and Russian (Mor)phonotactics", In.: Wrembel, M., Kiełkiewicz-Janowiak, A., & Gąsiorowski, P. (eds.) *Approaches to the Study of Sound Structure and Speech: Interdisciplinary Work in Honour of Katarzyna Dziubalska-Kołaczyk (1st ed.)*. Routledge.

Routledge Studies in Linguistics

APPROACHES TO THE STUDY OF SOUND STRUCTURE AND SPEECH

INTERDISCIPLINARY WORK IN HONOUR OF KATARZYNA DZIUBALSKA-KOŁACZYK

Edited by
Magdalena Wrembel, Agnieszka Kiełkiewicz-Janowiak and Piotr Gąsiorowski

ROUTLEDGE

Approaches to the Study of Sound Structure and Speech

This innovative work highlights interdisciplinary research on phonetics and phonology across multiple languages, building on the extensive body of work of Katarzyna Dziubalska-Kołaczyk on the study of sound structure and speech.

The book features concise contributions from both established and up-and-coming scholars who have worked with Katarzyna Dziubalska-Kołaczyk across a range of disciplinary fields toward broadening the scope of how sound structure and speech are studied and how phonological and phonetic research is conducted. Contributions bridge the gap between such fields as phonological theory, acoustic and articulatory phonetics, and morphology, but also includes perspectives from such areas as historical linguistics, which demonstrate the relevance of other linguistic areas of inquiry to empirical investigations in sound structure and speech. The volume also showcases the rich variety of methodologies employed in existing research, including corpus-based, diachronic, experimental, acoustic, and online approaches and showcases them at work, drawing from data from languages beyond the Anglocentric focus in existing research.

The collection reflects on Katarzyna Dziubalska-Kołaczyk's pioneering contributions to widening the study of sound structure and speech and reinforces the value of interdisciplinary perspectives in taking the field further, making this key reading for students and scholars in phonetics, phonology, sociolinguistics, psycholinguistics, and speech and language processing.

Magdalena Wrembel is University Professor and Head of Studies in the Faculty of English at Adam Mickiewicz University in Poznań, Poland. Her main research areas involve bilingualism and multilingualism, phonological acquisition of the third language, and language awareness. Her current work focuses on cross-linguistic influence and longitudinal development of L3 phonology.

Agnieszka Kiełkiewicz-Janowiak is University Professor in the Faculty of English at Adam Mickiewicz University in Poznań, Poland. She has done research and lectured on social dialectology, historical sociolinguistics, discourse analysis as well as language and gender issues. Her current research interests focus on life-span sociolinguistics and the discourse of ageing across cultures.

Piotr Gąsiorowski is University Professor in the Faculty of English at Adam Mickiewicz University in Poznań, Poland. His research interests include historical and evolutionary linguistics, theories of language change, dialectology, and phonetics and phonology. His current research work focuses on various aspects of Germanic and Indo-European reconstruction as well as Modern English prosody.

Routledge Studies in Linguistics

Perspectives from Systemic Functional Linguistics
Edited by Akila Sellami-Baklouti and Lise Fontaine

Time Series Analysis of Discourse
Method and Case Studies
Dennis Tay

Heart-and Soul-Like Constructs across Languages, Cultures, and Epochs
Edited by Bert Peeters

Systemic Functional Political Discourse Analysis
A Text-based Study
Eden Sum-hung Li, Percy Luen-tim Lui and Andy Ka-chun Fung

Systemic Functional Language Description: Making Meaning Matter
Edited by J.R. Martin and Y.J. Doran

Rarely Used Structures and Lesser-Studied Languages
Insights from the Margins
Emily Manetta

Externalization
Phonological Interpretations of Syntactic Objects
Yoshihito Dobashi

Approaches to the Study of Sound Structure and Speech
Interdisciplinary Work in Honour of Katarzyna Dziubalska-Kołaczyk
Edited by Magdalena Wrembel, Agnieszka Kiełkiewicz-Janowiak and Piotr Gąsiorowski

For more information about this series, please visit: www.routledge.com/Routledge-Studies-in-Linguistics/book-series/SE0719

Approaches to the Study of Sound Structure and Speech

Interdisciplinary Work in Honour of Katarzyna Dziubalska-Kołaczyk

Edited by Magdalena Wrembel, Agnieszka Kiełkiewicz-Janowiak and Piotr Gąsiorowski

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

First published 2020
by Routledge
52 Vanderbilt Avenue, New York, NY 10017

and by Routledge
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

Library of Congress Cataloging-in-Publication Data
A catalog record for this book has been requested

ISBN: 978-0-367-33760-5 (hbk)
ISBN: 978-0-429-32175-7 (ebk)

Typeset in Sabon
by Apex CoVantage, LLC

Contents

10 Main Differences Between German and Russian (Mor)phonotactics

A Corpus-Based Study

Wolfgang U. Dressler and Alona Kononenko-Szoszkiewicz

Introduction

The main aim of this study is to contrast phonotactics and morphonotactics of a major Germanic and a major Slavic language in regard to general typological properties of Germanic vs. Slavic languages. The criteria are similarities and dissimilarities of phonotactic and morphonotactic consonant clusters, their distribution in peripheral word positions, the impact of morphological vs. phonological properties on morphonotactic clusters, the distribution of preferred vs. dispreferred clusters (according to NAD criteria). In this way we hope to do justice to the pioneering contributions of Katarzyna Dziubalska-Kołaczyk to this area of linguistics and to our long-lasting cooperation since the late seventies.

The pertinent domain of our cooperation started with our joint proposal (Dressler & Dziubalska-Kołaczyk 2006, cf. Dressler, Dziubalska-Kołaczyk & Pestal 2010), followed by mutual consultation on work on Polish and English (Zydorowicz et al. 2015), German (Calderone et al. 2014; Celata et al. 2015; Freiberger 2014; Korecky-Kröll et al. 2014; Korecky-Kröll & Dressler 2015; Dressler & Kononenko 2019) and Slovak (Dressler et al. 2015), and by papers contrasting morphonotactics and phonotactics in Polish and English (Orzechowska 2009; Baumann & Kaźmierski 2016), Polish and German (Orzechowska & Wiese 2015; Wiese et al. 2017) and Slovak and German (Dressler et al. 2015).

The theoretical basis is a model of morphonology understood as an area of interaction between Natural Morphology and Natural Phonology (Dressler 1985, 1996). Its subpart morphonotactics is the area of interaction between phonotactics and morphotactics (Dressler & Dziubalska-Kołaczyk 2006; Zydorowicz et al. 2015). Since Natural Linguistics is a semiotically-based preference model (Dressler 1999; Dziubalska-Kołaczyk & Weckwerth 2002), we are going to study morphonotactic preferences of German and Russian, which are obviously typologically differentiated, thus this study has to be related to the subtheory of typological adequacy of both Natural Phonology (Dressler 1984) and Natural Morphology (Dressler & Kilani-Schoch 2017). Thus, Russian such as other Slavic languages is a more consonantal language than German and other Germanic languages, in terms of the number of both consonant phonemes and consonant clusters (cf. Dressler et al. 2015). In morphology, Russian is a more prefixing language than German. For phonotactic preferences we use the Beats-and-Binding Model (Dziubalska-Kołaczyk 2002, 2014). By taking into account the perceptual contrast between beats (vowels) and non-beats (consonants) it allows to evaluate cluster preferability and establish a hierarchy of the preferences of clusters from the most preferred (unmarked) to the least (marked). Perceptual contrast of the consonants is measured by means of the Net Auditory Distance principle (Dziubalska-Kołaczyk 2009).

Due to limits of space we are restricting our study to the most spectacular consonant word-initial and word-final clusters, triple and quadruple consonant clusters which contain at least two obstruents. All are in synchrony directly due to concatenative morphology, with the exception that all word-final quadruple Russian clusters arise in zero genitives of nouns with the final suffix *-stv+o*.

We limit ourselves to morphonotactic clusters, but include whether the same cluster is only morphonotactic or also phonotactic. We include both lemmas and inflectional forms, however exclude names and dialect words present in our data bases but not admitted in the standard.

Electronic Databases

For German, here presented by the Austrian standard, we used the Austrian Media Corpus (AMC), developed at the Austrian Academy of Science (cf. Ransmayr et al. 2017). It contains the words of all the Austrian print media of about the last 30 years, in sum about 40 million texts of various genres containing about 10 billion word tokens. It is linguistically annotated with morphosyntactic informations and lemmatized.

The data for Russian has been extracted for the needs of the project headed by Orzechowska et al. (2018) entitled: "The role of phonological features in phonotactics: A study on structure and learnability of consonant clusters in Slavic and Germanic languages" financed by the National Science Centre, Poland, under grant no. 2015/18/E/HS2/00066. About 37 thousand words were selected from The Great Russian-Polish Dictionary edited by "Wydawnictwo PWN" (Wawrzyńczyk et al. 2007) for further analysis.

Russian Word-Initial Clusters

All 13 quadruple consonant clusters consist of 3 obstruents followed by a sonorant /r/, /v/ or /l/. They are all (but 1) exclusively morphonotactic due to the presence of the prefixes /vz/, /vs/ and /v/ with the only exception /fstr-/ which is both morphonotactic in *f+stroit'* 'to build in' vs. synchronically phonotactic *vstreča* 'meeting'.

Cluster	Frequency in the corpus	Example	Translation
/vzbr-/	1	vz+bredat’	‘to come into one’s head’
/vzbl-/	1	vz +blesk	‘shine’
/vzgl-/	1	vz+gljad	‘look’
/vzgr-/	2	vz+grustnut’	‘to feel sad’
/vzdr-/	2	vz+dremnut’	‘to take a nap’
/fskl-/	2	vs+klokočenyj’	‘unkempt’
/vzdv-/	1	vz+dvoit	‘to double’
/fskr-/	3	vs+krikivat’	‘to scream’
/fspl-/	3	vs+plesk	‘splash’
/fspr-/	2	vs+prysnut’	‘to sprinkle’
/fstr-/	20	fs+trepat’	‘to dishevel’
/vsxl-/	1	vs+xlipyvat’	‘to sob’
/vsxr-/	1	vs+xrapyvat’	‘to snore’

Russian Word-Initial Triple Clusters

There are 33 exclusively morphonotactic word-initial triple clusters which can be divided into 4 groups according to the prefixes involved, for instance:

1. *v*-in adverbs, verbs as in *v+slepuju* ‘blindly’, *v+gnut’* ‘to bend’;
2. *vz*-in nouns or verbs as in *vz+glad* ‘look’, *vz+rastit’* ‘to nurture’;
3. *vs*-in nouns or nonfinite verbs *vs+xod* ‘rise’, *vs+xodit’* ‘to rise’;
4. *s*-in nouns and finite or nonfinite verbs *s+glaz* ‘an evil eye’, *s+družít'sja* ‘to become friends’.

The examples of triple consonant clusters are as follows:

Cluster	Frequency in the corpus	Example	Translation
/vbl-/	1	v+blizi	‘nearby’
/fkr-/	6	v+krutit’	to screw in’
/fpr-/	16	v+pravit’	‘to straighten’
/fsk-/	1	v+skok	‘a jump’
/fsl-/	3	v+sled	‘following’
/fsm-/	2	v+smotret’s’a	‘to peer’
/vbr-/	1	v+brosit’	‘to throw in’
/vgl-/	1	v+gladets’a	‘to gaze’
/vgn-/	1	v+gnut’	‘to bend’

/vgr-/	1	v+gryzats'a	'to gnaw'
/vdv-/	2	v+dvigat'	'to slide'
/fkl-/	18	v+klučit'	'to turn on'
/fst-/	13	v+stavit'	'to insert'
/ftr-/	3	v+troje	'three times as much'
/vzb-/	21	vz+badrivat'	'to cheer up'
/vzv-/	21	vz+valit'	'to charge'
/vzg-/	1	vz+gorje	'hill'
/vzd-/	6	vz+dox	'breath'
/vzl-/	10	vz+lom	'burglary'
/vzm-/	13	vz+max	'swing'
/vzn-/	1	vz+nos	'contribution'
/vzr-/	13	vz+rastit'	'to nurture'
/fsx-/	2	vs+xodit'	'to rise'
/fsp-/	30	vs+parit'	'to steam up'
/sgl-/	3	s+glaz	'an evil eye'
/sgn-/	1	s+gnit'	'to rotten'
/sgr-/	6	s+grebat'	'to shovel'
/sdv-/	5	s+dvig	'a shift'
/sxv-/	5	s+xvatit'	'to catch'
/sxl-/	4	s+xlestnutsja	'to clash'
/sxr-/	1	s+xryapat'	'to eat' colloquial
/zbl-/	5	s+blizit'	'to pull together'
/zdr-/	3	s+družít'sja	'to become friends'

There are 8 exclusively phonotactic consonant clusters which mainly occur in loan words such as *šplint* 'split pin', *škval* from English 'squall' with a rare case of turning a sibilant into a shibilant, or derived from Proto Slavic as in *smrad* 'a stench'.

Cluster	**Frequency in the corpus**	**Example**	**Translation**
/pxn-/	1	pxnut'	'to push'
/smr-/	1	smrad	'a stench'
/zdr-/	7	zdrávstvujte	'hello'
/sfr-/	1	sfragistika	'sphragistics'
/ʃkv-/	3	škval	'squall'
/ʃpl-/	1	šplint	'split pin'

/ʃpr-/	2	špric	'syringe'
/ʃtr-/	9	štraf	'a fine'

There are 9 ambiguous consonant clusters which occur in both morphonotactic and phonotactic consonant clusters:

Cluster	Frequency in the corpus	Example	Translation
/zbr-/	15	morph s+brosit' vs phon sbruja	'to reset', 'harness'
/skl-/	33	morph s+kleit' vs phon sklep	'to glue', 'a tomb'
/skv-/	17	morph s+kvasit' vs phon skwaz̨ina	'to ferment', 'well'
/skr-/	55	morph s+kryt' vs phon skrip'	'to hide', 'creak'
/spl-/	41	morph s+plesti vs phon English loanwords splin	'to weave', 'spleen'
/spr-/	37	morph s+pryatat' vs phon loanword 'sprint'	'to hide'
/str-/	177	morph s+trusit' vs phon strana	'to quail', 'a country'
/stv-/	5	morph s+tvorožitsja vs phon 'stvol'	'to curdle', 'a trunk'
/stl-/	2	morph s+tlet' vs phon stlat'	'to smolder', to spread'

Most of the word-initial clusters are diachronically due to deletion of the reconstructed Protoslavic unstressed short high vowels, which is also the origin of Slavic vowelless consonant prefixes. Thus a change of phonotactics is responsible for the great quantity and variety of word-initial complex consonant clusters. The great morphological productivity of prefixation adds up to the number of clusters and their type frequency and to the existence of morphonotactic quadruple clusters, which are cross-linguistically a very marked category.

Word-Final Consonant Clusters in Russian

In Russian, word-final position consonant clusters are not as numerous as word-initially. Thus, there are only 3 Russian quadruple word-final clusters which occur only in nouns in Gen. case Pl. ending in the suffix *-stvo*, thus all exclusively morphonotactic:

Cluster	Frequency in the corpus	Example	Translation
/-rstf/	61	mytar+stfv	'hardship'
/-fstf/	3	graf+stv	'county'
/-pstf/	7	poxab+stv	'obscenity'

Russian Word-Final Triple Clusters

In comparison to quadruple consonant clusters, word-final triple clusters are mainly phonotactic due to their origin. There are 3 word-final clusters which occur only in nouns, mainly in loan words, either phonotactic in Nom. case Sg. as in /-str/: in loan words *ministr* 'minister', *semestr* 'semester'; /-ktr/: *spektr* 'spectrum' or morphonotactic in Gen. Pl. in /-stf/: (*zodčestv* 'of architectures') of nouns ending in the suffix—*stvo*.

(Dis)preferences According to NAD

The NAD calculator has been designed by Dziubalska-Kołaczyk et al. (2007, 2009, 2014) for measuring the auditory distances between the neighboring phonemes. It allows for measuring the preferability of the cluster according to its position in a word (initial, medial or final, respectively) and to build up the hierarchy of preferability of clusters from the most to the least preferred. The parameters of the calculator include Manner and Place of Articulation as well as an obstruent-sonorant distinction. This NAD calculator functions for maximally triple consonant clusters.

The hierarchy of preferability values for word-initial triple consonant clusters in Russian from the most to the least preferred:

IPA transcription	CV structure	NAD (VC)	NAD (C1C2)	NAD (C2C3)	NAD (CV)	NAD product	Preferred cluster?	Phon or morph, or both
fprV	CCCV	-	1.5	5.3	2	3.55	Yes	Morph
vbrV	CCCV	-	1.5	5.3	2	3.55	Yes	Morph
sbrV	CCCV	-	2	5.3	2	3.3	Yes	Both
sprV	CCCV	-	2	5.3	2	3.3	Yes	Both
ʂprV	CCCV	-	2.3	5.3	2	3.15	Yes	Phon
zgrV	CCCV	-	2.5	5.2	2	2.95	Yes	Morph
skrV	CCCV	-	2.5	5.2	2	2.95	Yes	Both
zdrV	CCCV	-	1	4.3	2	2.8	Yes	Morph
zdrV	CCCV	-	1	4.3	2	2.8	Yes	Phon
strV	CCCV	-	1	4.3	2	2.8	Yes	Both
fkrV	CCCV	-	3	5.2	2	2.7	Yes	Morph
vgrV	CCCV	-	3	5.2	2	2.7	Yes	Morph
ʂtrV	CCCV	-	1.3	4.3	2	2.65	Yes	Phon
ftrV	CCCV	-	1.5	4.3	2	2.55	Yes	Morph
sfrV	CCCV	-	0.5	3.8	2	2.55	Yes	Phon
vblV	CCCV	-	1.5	4.5	2.5	2.5	Yes	Morph
zglV	CCCV	-	2.5	5	2.5	2.5	Yes	Morph
sklV	CCCV	-	2.5	5	2.5	2.5	Yes	Both
sxrV	CCCV	-	1.5	4.2	2	2.45	Yes	Morph
vglV	CCCV	-	3	5	2.5	2.25	Yes	Morph

fklV	CCCV	-	3	5	2.5	2.25	Yes	Morph
zblV	CCCV	-	2	4.5	2.5	2.25	Yes	Morph
splV	CCCV	-	2	4.5	2.5	2.25	Yes	Both
ʂplV	CCCV	-	2.3	4.5	2.5	2.1	Yes	Phon
vzrV	CCCV	-	0.5	3.3	2	2.05	Yes	Morph
sxlV	CCCV	-	1.5	4	2.5	2	Yes	Morph
zgnV	CCCV	-	2.5	4.5	3	1.75	Yes	Morph
stlV	CCCV	-	1	3.5	2.5	1.75	Yes	Both
vgnV	CCCV	-	3	4.5	3	1.5	Yes	Morph
fsmV	CCCV	-	0.5	3	3	1.25	Yes	Morph
vzmV	CCCV	-	0.5	3	3	1.25	Yes	Morph
fslV	CCCV	-	0.5	2.5	2.5	1	Yes	Morph
vzlV	CCCV	-	0.5	2.5	2.5	1	Yes	Morph
vznV	CCCV	-	0.5	2	3	0.25	No	Morph
pxnV	CCCV	-	3.5	3.5	3	0.25	No	Phon
smrV	CCCV	-	3	2.3	2	-0.2	No	Phon
ʂkvV	CCCV	-	2.2	3	5	-0.6	No	Phon
fskV	CCCV	-	0.5	2.5	6	-0.75	No	Morph
vzgV	CCCV	-	0.5	2.5	6	-0.75	No	Morph
skvV	CCCV	-	2.5	3	5	-0.75	No	Both
vzbV	CCCV	-	0.5	2	6	-1.25	No	Morph
fsxV	CCCV	-	0.5	1.5	5	-1.25	No	Morph
fspV	CCCV	-	0.5	2	6	-1.25	No	Morph
sxvV	CCCV	-	1.5	2	5	-1.25	No	Morph
sdvV	CCCV	-	1	1.5	5	-1.5	No	Morph
stvV	CCCV	-	1	1.5	5	-1.5	No	Both
vdvV	CCCV	-	1.5	1.5	5	-1.75	No	Morph
fstV	CCCV	-	0.5	1	6	-2.25	No	Morph
vzvV	CCCV	-	0.5	0.5	5	-2.25	No	Morph
vzdV	CCCV	-	0.5	1	6	-2.25	No	Morph

The hierarchy of preferability values for word-final triple consonant clusters in Russian from the most to the least preferred:

IPA transcription	CV structure	NAD (VC)	NAD (C1C2)	NAD (C2C3)	NAD (CV)	NAD product	Preferred cluster?	Phon or morph, or both
Vstr	CCCV	5	1	4.3	-	-3.65	No	Phon
Vktr	CCCV	6	1.5	4.3	-	-3.65	No	Phon
Vstf	CCCV	5	1	1.5	-	-2.25	No	Morph

Thus, among word-initial clusters, 21 only morphonotactic ones are preferred, 12 dispreferred, among only phonotactic ones, 5 are preferred and 3 dispreferred, and among both morphonotactic and phonotactic clusters, 7 are preferred and 2 dispreferred. Hence, Russian also prefers universally unmarked consonant clusters, both among morphonotactic and phonotactic clusters, which supports our claim of a big amount of harmony between phonotactic and morphonotactic clusters. But the claim (Dressler & Dziubalska-Kołaczyk 2006) that morphonotactic clusters are generally more dispreferred than phonotactic ones has been disproved. It must be reduced to the claim that morphonotactic clusters can be more complex in number of member consonants and in number of types than phonotactic ones. At least this is true for Russian.

German Word-Initial Clusters

Since Standard German has no vowelless consonantal prefixes, all word-initial consonant clusters are phonotactic. There are no quadruple clusters and the triple clusters consist of /ʃ/plus /tr, pr, pl/ and of old and new loan-words with /s/ plus /tr, pr, kr, kl, pl, kv/. Thus, the small quantity of phonotactic clusters is not increased by morphonotactics.

German Word-Final Clusters

All 25 word-final quadruple clusters consist first of a sonorant (only twice a fricative) and then normally of 3 obstruents. There are also 4 cases of 2 sonorants plus 2 obstruents. The two last obstruents are always /st/, with 4 single types of genitive singulars in 4 clusters ending in /-st͡s/. All are either only morphonotactic (17) or by a strong default (4), are nearly only verb forms and represent mainly the 2nd Sg. (+*st*), sometimes also the 3rd Sg. or the past participle (+*t*), plus 2 opaque superlatives. In parenthesis the number of types in the AMC is given (prefixed verbs, including particle verbs, are not counted; not all potential 2nd Sg. forms occur in the AMC):

Cluster	Frequency in the corpus	Example	Translation
/-lkst/	5	melk+st, verfolg+st	'you milk', 'you persecute'
/-rkst	30	merk+st, borg+st, verkorks+t	'you notice', 'you borrow', 'messed up'
/-mpst/	11	pump+st, plumps+t	'you pump', '(s)he flops' = plumps+st 'you flop' (with obligatory degemination of /s+s/),
/-mp͡fst/	10	kämpf+st	'you fight'
/-lʃst/	2	fälsch+st	'you falsify'
/-nʃst/	3	wünsch+st	'you want'
/-nt͡ʃst/	3	plantsch+st, English loan-words launch+st, lunch+st,	'you splash'
/-lfst/	3	hilf+st	'you help'

/-rfst/	65	darf+st, nerv+st	'you may', 'you enervate'
/-rmst/	29	form+st	'you form'
/-lmst/	8	film+st	'you film'
/-rnst/	8	morph lern+st vs phon ernst	'you learn', 'earnest' (and its homophonus noun)
/-lxst/	2	strolch+st	'you roam about'
/-rxst/	11	schnarch+st	'you snore'
/-fts͜st/	1	seufz+st, normally the /s/ is fused with the preceding affricate	'you sigh'
/-xts͜st/	3	ächz+st (same fusion),	'you groan'
/-rts͜st/	2	stürz+st (same fusion),	'you fall'
/-lts͜st/	2	salz+st (same fusion).	'you salt'
/-ŋksts͜/	1	Hengst+s	'stallion' (masc.)
/-rpsts͜/	1	Herbst+s	'autumn' (masc.)
/-lpsts͜/	1	Selbst+s	'the self' (neuter)
/-rnsts͜/	1	Ernst+s	'earnestness' (masc.)
/-ŋkst/	37	denk+st, sing+st; morphosemantically somewhat opaque superlatives jüng+st; läng+st vs. phonotactic only Angst, Hengst	'you think', variant pronunciation of /-ngst/: 'you sing', 'recently', 'for a long time''fear', 'stallion'
/-rpst/	204	morph stirb+st vs. phonotactic only Herbst	'autumn' (plus its many compounds)
/-lpst/	5	stülp+st vs. selb(+)st	'you turn up (the collar)', 'oneself' with a fossile suffix

Triple clusters are obviously more numerous (additionally especially in types and tokens) and varied than quadruple clusters. Nearly all of the clusters ending in *-t* start with a sonorant, the two final obstruents are / st/ and /ft/. They are 25 exclusively morphonotactic ones, 10 as defaults, 1 phonotactic by default. The additional morphological forms are transparent superlatives, ordinal numbers, (mostly opaque) deverbal nouns and the circumfixation *ver+wanz+t* 'bug-infested'. The exclusively morphonotactic ones (without giving examples) are: /-xst, -xts͜t, -fst, -mst, -ʃst, -pf͜st, -t͜ʃst, -fts͜t, -xts͜t, -nts͜t, -lts͜t, -lft, -lxt, -lʃt, -mʃt, -rt͜ʃt, -nʃt, -pʃt, -rʃt, -nxt, -nt͜ʃt, -lkt, -mpf͜t, -mpt, -rpt, -lpt/. Morphonotactic by default with clearly phonotactic exceptions are: /-rts͜t, -rst, -pst, -rkt, -ŋkt, -rxt, -rft, -kst/. Most phonotactic exceptions are diachronically (and maybe for some speakers still synchronically) morphonotactic for /-lst/, e.g., in *Ge+schwul+st* 'tumor' from *schwell+en* 'to swell', and /-nst/, e.g., *Kunst* 'art' from *könn+en* 'to be able'. The only cluster which is phonotactic by default is /-nft/, as in *sanft* 'soft' vs. the only morphonotactic type *fünf+t* 'fifth'.

In contrast to quadruple clusters, triple clusters may end also in *-s* or affricate *-t͡s* due to stem-final dental (*-t, -d, -t͡s*) merging with a suffix *-s*. All obstruents may precede the *-s*, and the first consonant is mostly a sonorant, but may be also a fricative or affricate or stop. The suffix *-s* may signal the plural, the genitive singular or an adverb. The exclusively morphonotactic clusters are /-rps, -rfs, -lfs, -nfs, -lks; -ŋkt͡s, -rst͡s, -rpt͡s, -lst͡s, -nst͡s/ plus the masculine genitives /-rt͡st͡s -pst͡s, -rkt͡s, -kst͡s/, each represented by a single type; the genitive *des Arzt+s* is pronounced in formal speech with a double affricate, thus /-rt͡st͡s/. The clusters /-rks, -lps, -ŋks/ have each just 1 phonotactic exception: *Murks* 'botch', *Rülps* 'belch', *Sphinx*. Morphonotactic by default are /-rps, -mps/. Hence, all German clusters are only or mostly morphonotactic (with 1 exception of a phonotactic default for a *-t*-final cluster).

The hierarchy of preferability values for word-final triple consonant clusters in German ending in *-t* from the most to the least preferred:

IPA transcription	**CV structure**	**NAD (C1C2)**	**NAD (C2C3)**	**NAD (CV)**	**NAD product**	**Preferred cluster?**	**Phon or morph, or both**
Vrpt	VCCC	2	6.6	1	5.1	Yes	Both
Vrt͡st	VCCC	2	5.1	0.5	3.85	Yes	Both
Vrft	VCCC	2	5.1	1.5	3.35	Yes	Both
Vrst	VCCC	2	4.6	1	3.1	Yes	Both
Vlkt	VCCC	2.5	4.8	1.3	2.9	Yes	Morph
Vlpt	VCCC	2.5	4.5	1	2.75	Yes	Both
Vrkt	VCCC	2	4.3	1.3	2.65	Yes	Both
Vrʃt	VCCC	2	4.1	1.5	2.35	Yes	Both
Vlxt	VCCC	2.5	5.5	4	2.25	Yes	Both
Vnkt	VCCC	3	4.3	1.3	2.15	Yes	Morph
Vnxt	VCCC	3	5	4	1.5	Yes	Morph
Vltst	VCCC	2.5	3	0.5	1.5	Yes	Morph
Vrxt	VCCC	2	4.4	4	1.4	Yes	Morph
Vmʃt	VCCC	3	3.5	1.5	1.25	Yes	Morph
Vmp͡ft	VCCC	3	3	1	1	Yes	Morph
Vlft	VCCC	2.5	3	1.5	1	Yes	Morph
Vmst	VCCC	3	3	1	1	Yes	Morph
Vmpt	VCCC	3	3	1	1	Yes	Both
Vlʃt	VCCC	2.5	3	1.5	1	Yes	Morph
Vt͡st	VCCC	5	3.5	0.5	0.75	No	Morph
Vnt͡st	VCCC	3	2.5	0.5	0.75	No	Morph

Vlst	VCCC	2.5	2.5	1	0.75	Yes	Both
Vnft	VCCC	3	2.5	1.5	0.25	No	Morph
Vnʃt	VCCC	3	2.5	1.5	0.25	No	Morph
Vxst	VCCC	5	3	1	0	No	Morph
Vnst	VCCC	3	2	1	0	No	Morph
Vkst	VCCC	6	2.3	1	-1.2	No	Both
Vpʃt	VCCC	6	2.5	1.5	-1.25	No	Morph
Vpst	VCCC	6	2	1	-1.5	No	Morph
Vft͡st	VCCC	5	1	0.5	-1.75	No	Morph
Vp͡fst	VCCC	5.5	1	1	-2.25	No	Both
Vfst	VCCC	5	0.5	1	-2.5	No	Both
Vʃst	VCCC	5	0.5	1	-2.5	No	Morph

Among the 33 word-final consonant clusters 20 clusters are preferred and 13 dispreferred, which contrasts with the much greater percentage of preferred clusters in Russian. Among the exclusively morphonotactic clusters there is an equal number (10) of preferred and dispreferred ones. Among the both morphonotactic and phonotactic ones the preferred clusters are the large majority (10 vs. 3). This represents only a weak support for the claim that morphonotactic clusters are more dispreferred than phonotactic ones. However, there is more support for our modified version that morphonotactic clusters are more responsible for the complexity of clusters than phonotactic clusters.

The hierarchy of preferability values for word-final triple consonant clusters in German ending in *-s* or *-t͡s* from the most to the least preferred:

IPA transcription	**CV structure**	**NAD (VC)**	**NAD (C1C2)**	**NAD (C2C3)**	**NAD product**	**Preferred cluster?**	**Phon or morph, or both**
Vrpt͡s	VCCC	2	6.6	1.5	4.85	Yes	Morph
Vrps	VCCC	2	6.6	2	4.6	Yes	Both
Vrt͡st͡s	VCCC	2	5.1	0	4.1	Yes	Morph
Vrfs	VCCC	2	5.1	0.5	3.85	Yes	Both
Vrst͡s	VCCC	2	4.6	0.5	3.35	Yes	Morph
Vlks	VCCC	2.5	4.8	2.3	2.4	Yes	Morph
Vrkt͡s	VCCC	2	4.3	1.8	2.4	Yes	Morph
Vrkt͡s	VCCC	2	4.3	1.8	2.4	Yes	Morph
Vlps	VCCC	2.5	4.5	2	2.25	Yes	Both
Vrks	VCCC	2	4.3	2.3	2.15	Yes	Morph
Vlfs	VCCC	2.5	3	0.5	1.5	Yes	Morph
Vlst͡s	VCCC	2.5	2.5	0.5	1	Yes	Morph

Vnfs	VCCC	3	2.5	0.5	0.75	No	Morph
Vŋkt͡s	VCCC	3	3	1.8	0.6	Yes	Morph
Vmps	VCCC	3	3	2	0.5	Yes	Morph
Vŋks	VCCC	3	3	2.3	0.35	Yes	Both
Vnst͡s	VCCC	3	2	0.5	0.25	No	Morph
Vkst͡s	VCCC	6	2.3	0.5	-0.95	No	Morph
Vpst͡s	VCCC	6	2	0.5	-1.25	No	Morph

There are 19 word-final clusters. Ten affricate-final clusters are exclusively morphonotactic among which 7 clusters are preferred and 3 dispreferred; among 9 clusters ending on *-s*: only *-nfs* is morphonotactic and dispreferred, while the rest 8 clusters are preferred. Thus, *-s* final clusters are much more preferable than *-t* final clusters, although they are only or mostly morphonotactic.

Conclusions

In this contrastive study of German and Russian morphonotactic and phonotactic triple and quadruple consonant clusters in both word-initial and word-final position we have found important typological differences:

1. German prefers complex clusters word-finally, Russian (like Polish and Slovak) word-initially.
2. As to morphonotactic clusters, the reason is the non-existence of purely consonantal prefixes and thus of word-initial morphonotactic consonant clusters in standard German vs. the non-existence of purely consonantal suffixes which attach to a preceding root-final consonant without vowel insertion in Russian, which leads to a rarity of word-final morphonotactic consonant clusters: those which exist are only due to zero plural genitives.
3. Partially due to 2., the relative frequency of morphonotactic vs. phonotactic clusters is different.
4. But these distributional differences between German and Russian morphonotactic clusters harmonizes at least tendentially with the distribution of phonotactic clusters. Diachronically this difference in the distribution of phonotactic clusters is due to the German loss of (always unstressed) schwa in word-final syllables, and to the Old Slavic rhythmic loss of unstressed ultrashort consonants throughout the word, plus the loss of many word-final consonants. These phonological changes have also had an impact on the morphologies of both languages.
5. The distribution of preferred vs. dispreferred morphonotactic clusters (based on NAD) varies in German vs. Russian.
6. The amount of morphonotactic clusters in Russian is greater due to both Russian being more of a consonantal language and having a

richer inflectional and derivational morphology than German, whereby the latter explains why the majority of word-initial consonant clusters of Russian is morphonotactic.

7. The majority of morphonotactic clusters in Russian are preferred, which goes against the claim that consonantal languages are expected to have more dispreferred clusters and that morphonotactics renders clusters to be dispreferred.

References

Baumann, A. & Kaźmierski, K. 2016. A dynamical-systems approach to the evolution of morphonotactic and lexical consonant clusters in English and Polish. *Yearbook of the Poznań Linguistic Meeting* 2. 115–139.

Calderone, B., C. Celata, K. Korecky-Kröll & W. U. Dressler. 2014. A computational approach to morphonotactics: Evidence from German. *Language Sciences* 46. 59–70.

Celata, C., K. Korecky-Kröll, I. Ricci & W. U. Dressler. 2015. Phonotactic processing and morpheme boundaries: Word-final /Cst/ clusters in German. *Italian Journal of Linguistics* 27. 85–110.

Dressler, W. U. 1984. Explaining Natural Phonology. *Phonology Yearbook* 1. 29–51.

Dressler, W. U. 1985. *Morphonology*. Ann Arbor: Karoma Publishers.

Dressler, W. U. 1996. A functionalist semiotic model of morphonology. In R. Singh (ed.), *Trubetzkoy's orphan*, 67–83. Amsterdam: Benjamins.

Dressler, W. U. 1999. On a semiotic theory of preferences in language. *Peirce Seminar Papers* 4. 389–415.

Dressler, W. U. & K. Dziubalska-Kołaczyk. 2006. Proposing morphonotactics. *Italian Journal of Linguistics* 18. 249–266.

Dressler, W. U., K. Dziubalska-Kołaczyk & L. Pestal. 2010. Change and variation in morphonotactics. *Folia Linguistica Historica* 31. 51–67.

Dressler, W. U., M. Hliničanová, M. Ďurčo, K. Mörth & K. Korecky-Kröll. 2015. Phonotactic vs. morphonotactic obstruent clusters in Slovak and German. *Italian Journal of Linguistics* 27. 45–60.

Dressler, W. U. & M. Kilani-Schoch. 2017. Natural Morphology. In A. Hippisley & G. Stump (eds.), *The Cambridge handbook of morphology*, 356–389. Cambridge: Cambridge University Press.

Dressler, W. U. & A. Kononenko. 2019 [submitted]. German phonotactic vs. morphonotactic obstruent clusters: A corpuslinguistic analysis. In W. U. Dressler, S. Sommer-Lolei, K. Korecky-Kröll & B. Calderone (eds.), *Experimental and acquisitional approaches to morphonotactics*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Dziubalska-Kołaczyk, K. 2002. *Beats-and-binding phonology*. Frankfurt: Peter Lang.

Dziubalska-Kołaczyk, K. 2009. NP extension: B&B phonotactics. *Poznań Studies in Contemporary Linguistics* 45. 55–71.

Dziubalska-Kołaczyk, K. 2014. Explaining phonotactics using NAD. *Language Sciences* 46. 6–17.

Dziubalska-Kołaczyk, K. & G. Krynicki. 2007. Phonotactic preferences in Polish, English and German: Quantitative perspective. Paper presented at the 38th Poznań Linguistic Meeting, 13–16 September.

Dziubalska-Kołaczyk, K. & J. Weckwerth (eds.). 2002. *Future challenges for natural linguistics*. Munich: Lincom.

Freiberger, E. M. 2014. Der Erwerb von Konsonantenclustern im Deutschen: Eine Untersuchung longitudinaler Spontansprachkorpora. *Slowakische Zeitschrift für Germanistik* 6. 5–23.

Korecky-Kröll, K. & W. U. Dressler. 2015. (Mor)phonotactics in high vs. low SES children. In K. Dziubalska-Kołaczyk & J. Weckwerth (eds.), *Rajendra Singh: In memoriam*, 25–40. Poznań: Adam Mickiewicz University Press.

Korecky-Kröll, K., W. U. Dressler, E. M. Freiberger, E. Reinisch, K. Mörth & G. Libben. 2014. Morphonotactic and phonotactic processing in German-speaking adults. *Language Sciences* 46. 48–58.

Orzechowska, P. 2009. *Morphonotactics in English and Polish: A dictionary and corpus-based study of word-final consonant clusters*. Ph.D. dissertation. Poznań: Adam Mickiewicz University.

Orzechowska, P., J. Mołczanow & M. Jankowski. 2018. Struktura sylaby a akcent wyrazowy w języku rosyjskim. *Studia et Documenta Slavica* 1–2(5–6). 79–91.

Orzechowska, P. & R. Wiese. 2015. Preferences and variation in phonotactics: A multi-dimensional evaluation of German and Polish. *Folia Linguistica* 49. 439–486. doi: 10.1515/flin-2015–0016.

Ransmayr, J., M. Karlheinz & M. Durco. 2017. AMC (Austrian Media Corpus)—Korpusbasierte Forschungen zum österreichischen Deutsch. In C. Resch & W. U. Dressler (eds.), *Digitale Methoden der Korpusforschung in Österreich*, 27–38. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Wawrzyńczyk, J., M. Kuratczyk, E. Małek & H. Bartwicka. 2007. *Wielki słownik rosyjsko-polski*. Warsaw: Wydawnictwo Naukowe PWN.

Wiese, R., P. Orzechowska, P. M. Alday & C. Ulbrich. 2017. Structural principles or frequency of use? An ERP experiment on the learnability of Polish consonant clusters. *Frontiers in Psychology: Auditory Cognitive Neuroscience* 7. 2005.

Zydorowicz, P., P. Orzechowska, M. Jankowski, K. Dziubalska-Kołaczyk, P. Wierzchoń & D. Pietrala. 2015. *Phonotactics and morphonotactics of Polish and English: Theory, description, tools and applications*. Poznań: Adam Mickiewicz University Press.

Appendix 9. Research article 3: Dressler & Kononenko-Szoszkiewicz, 2021)

Dressler, Wolfgang U. and Alona Kononenko-Szoszkiewicz, 2021, “German phonotactic vs. morphonotactic obstruent clusters: a corpus linguistic analysis “, In.: W. U. Dressler, B. Calderone, S. Sommer-Lolei, K. Korecky-Kröll (eds.) *Experimental, Acquisitional and Corpus linguistic Approaches to the Study of Morphonotactics*, Austrian Academy of Sciences Press.

WOLFGANG U. DRESSLER – BASILIO CALDERONE –
SABINE SOMMER-LOLEI – KATHARINA KORECKY-KRÖLL (EDS.)

EXPERIMENTAL, ACQUISITIONAL AND CORPUS LINGUISTIC APPROACHES TO THE STUDY OF MORPHONOTACTICS

AUSTRIAN
ACADEMY
OF SCIENCES
PRESS

WOLFGANG U. DRESSLER – BASILIO CALDERONE –
SABINE SOMMER-LOLEI – KATHARINA KORECKY-KRÖLL (EDS.)

EXPERIMENTAL, ACQUISITIONAL AND CORPUS LINGUISTIC APPROACHES TO THE STUDY OF MORPHONOTACTICS

ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN
PHILOSOPHISCH-HISTORISCHE KLASSE
SITZUNGSBERICHTE, 915. BAND

VERÖFFENTLICHUNGEN ZUR LINGUISTIK
UND KOMMUNIKATIONSFORSCHUNG

BAND 32

HERAUSGEGEBEN VON
WOLFGANG U. DRESSLER

Experimental, Acquisitional and Corpus linguistic Approaches to the Study of Morphonotactics

edited by

Wolfgang U. Dressler
Basilio Calderone
Sabine Sommer-Lolei
Katharina Korecky-Kröll

AUSTRIAN ACADEMY OF SCIENCES PRESS

Printed with the support of the
Austrian Science Fund (FWF): PUB 862-Z

FWF Der Wissenschaftsfonds.

This publication was subject to international and anonymous peer review.
Peer review is an essential part of the Austrian Academy of Sciences Press evaluation
process. Before any book can be accepted for publication, it is assessed by international
specialists and ultimately must be approved by the Austrian Academy of Sciences
Publication Committee.

The paper used in this publication is DIN EN ISO 9706 certified and meets
the requirements for permanent archiving of written cultural property.

ISBN 978-3-7001-8714-1

Layout: Andrea Sulzgruber, Vienna
Print: Prime Rate, Budapest
https://epub.oeaw.ac.at/8714-1
https://verlag.oeaw.ac.at
Made in Europe

TABLE OF CONTENTS

I. German phonotactic vs. morphonotactic obstruent clusters: a corpus linguistic analysis

Wolfgang U. Dressler[1,2]
Alona Kononenko-szoszkiewicz[1]

1. INTRODUCTION

1.1. Aims

In this contribution we provide for the first time a typological characterology (in the sense of Mathesius 1928; Lang & Zifonun 1996) of the morphonotactics vs. phonotactics of a single language, compared to contrastive studies such as Dressler et al. (2015) on German vs. Slovak and Zydorowicz et al. (2016) on Polish vs. English. We focus on word-initial and word-final positions (cf. section 4) and on triple consonant clusters (excluding glides) containing two obstruents, because these are more typical for German than for many other languages. We approach them in terms of an interaction between Natural Phonology and Natural Morphology and the Beats-and-Binding phonotactics of Dziubalska-Kołaczyk (2009). We limit our investigation to standard vocabulary and exclude onomastics, because it contains clusters that do not occur in standard vocabulary, such as *gm-* in many place names (*Gmünd, Gmunden* etc.).

With regard to phonological typology, German, like other Germanic languages, is a rather consonantal language in respect of the relative amount of its consonantal inventory and its variety and complexity of consonant clusters (cf. Maddieson 2006, 2013; Donohue et al. 2013), although – in contrast to several Slavic languages, for example – German has syllabic sonorants only in an unstressed position in casual speech. German has several voiceless affricates, among the typologically rather rare ones the labial-labiodental /p͜f/ (Luschützky 1992). German is richer in consonant clusters word-finally than word-initially, in contrast to most Romance and many other non-Germanic Indo-European languages. Phonological typology, though discussed at least since Trubetzkoy (1939),

---

[1] Austrian Centre for Digital Humanities and Cultural Heritage (ACDH-CH) of the Austrian Academy of Sciences, Vienna.
[2] University of Vienna.

has focused on the characteristics of phonemes, phoneme oppositions and phoneme inventories. If phonotactics has been treated at all, then it is in terms of syllable structures. Even the recent publications of Hyman (2007), Blevins (2007) and Hyman and Plank (2018) mention consonant clusters at most in passing and never discuss triple or quadruple clusters (for contrastive studies of German, see section 1.6). This lacuna may be due to phonological typologists not working with large electronic corpora, which we do for German in this contribution.

In continuation of previous theoretical and contrastive work (Dressler & Dziubalska-Kołaczyk 2006; Dressler, Dziubalska-Kołaczyk & Pestal 2010; Korecky-Kröll et al. 2014) we are going to characterize German patterns of consonantal morphonotactics vs. phonotactics from a phonological, morphological, typological and corpus linguistic perspective.

We investigate prototypical rather than non-prototypical cases of morphonotactics, i.e. the prototypical case of merely concatenative shapes of morpheme combinations, particularly when they differ from the phonotactics of lexical roots and morphemes and thus signal morpheme boundaries, as in English *seem-ed* /si:m-d/ (i.e. there is no lexical final [-md] cluster in English). The non-prototypical case of morphological combinations resulting in vowel deletion is marginal in German, e.g. in *Risiko* ‘risk’, adj. *risk-ant* ‘risky’ (in contrast to the regular case of schwa deletion, more in section 4).

1.2. Phonotactics vs. morphonotactics

Morphonotactic clusters differ from phonotactic ones through the interaction of morphotactics with phonotactics (Dressler & Dziubalska-Kołaczyk 2006; Calderone, Celata & Laks 2014; Zydorowicz et al. 2016). More specifically, morphonotactic clusters are either due to the addition of a further morpheme, an affix in the case of derivational morphology or another lexical morpheme in the case of compounding, or due to a subtractive morphotactic operation which leads to vowel deletion, as in Ger. *silbr-ig* ‘silvery’ from *Silber* ‘silver’ (more in section 4.2).

Because of this interaction between morphology and phonology, it has been claimed (Dressler & Dziubalska-Kołaczyk 2006: 19–20) that in general morphonotactic clusters are less preferred than phonotactic ones. This contrasts with the Strong Morphonotactic Hypothesis (Dressler & Dziubalska-Kołaczyk 2006; Dressler et al. 2010), which states that in processing and first language acquisition the interaction of morphology with phonotactics facilitates both processing and acquisition. A further claim on the interaction between morphology and phonology has been made by Shosted (2006), who has found a (statistically insignificant) trend of a positive correlation between complexity in the syllable structure and morphological complexity. It would be worth separating phonological and morphonotactic clusters, because only complex morphonotactics should correlate with morphological complexity.

In order to define the level of deviation of morphonotactic (i.e. morphologically and phonologically motivated) consonant clusters from purely phonotactic (i.e. merely phonologically motivated) ones in German, we have applied the gradual scale proposed by Dressler and Dziubalska-Kołaczyk (2006). These are clusters such as the following English ones:

1) Clusters which are always morphologically motivated, i.e. never occur in monomorphemic words (cf. Dressler 1985: 220 f.). To this group belongs a consonant cluster /-md/ which always occurs in past participles due to concatenation of a sonorant with the suffix, as in *seem-ed, claim-ed.* Other examples of this group are the word-final consonant clusters /-fs, -vz/ as in *laughs, loves, wife's, wives*, which occur only in plurals, third person singular present forms and in Saxon genitives.

2) Clusters, which are morphologically motivated as a strong default, i.e. which are paralleled by very few exceptions of a morphologically unmotivated nature. For instance, the cluster /ts/ in most cases occurs across word boundaries, as in *lets*, *meets,* but also in morphologically simple words as in *quartz*, *hertz*. Moreover, in English a strong default is present in a cluster /-ps/ as in *steps*, *keeps*, except the borrowings from Latin such as *apse*, *lapse*, and *glimpse*.

3) Clusters, which are morphologically motivated as a weak default, i.e. which are paralleled by more exceptions of a morphologically unmotivated nature. An example is the consonant cluster /-ks/, which is always morphonotactic in the third person singular verb endings and in plurals as in *speaks*, *oaks*, and a phonotactic cluster related to the spelling <x> as in *fox, mix.*

4) Clusters, whose minority is morphologically motivated, i.e. which are quite normal phonotactic clusters and may also have some morphological motivation. To this group belongs the cluster /-nd/ that occurs across morpheme boundaries in past-tense verbs or past participles as in *grinned, tanned*. Moreover, as a phonotactic cluster, it is present in a number of words such as *hand*, *land*, *around*.

5) Clusters which are only phonotactic, thus never divided by a morpheme boundary, such as /rf, sk/, as in *turf, ask*.

The theoretical background of our contribution is Natural Phonology and Morphology (cf. Dressler 1984; Dziubalska-Kołaczyk & Weckwerth 2002; Dziubalska-Kołaczyk 2009; Kilani-Schoch & Dressler 2005; Dressler & Kilani-Schoch 2017), as well as morphonology (Dressler 1985, 1996a,b), of which morphonotactics is a part (Dressler & Dziubalska-Kołaczyk 2006). This approach not only strives towards descriptive und explanatory adequacy but also towards guaranteeing, at least partially, the psychological reality of the linguistic constructs. This demands a psycholinguistic perspective (cf. Korecky-Kröll et al. 2014 and Sommer-Lolei et al. this volume). In usage-based linguistic and psycholinguistic approaches (Bybee 2001; Bauer 2001; Tomasello 2003), it is often claimed that token frequency is important only for the question of storage (which is not an issue here), whereas only type frequency and the discrepancy between high type frequency and low token frequency is relevant for the productivity and profitability of patterns (cf. Du & Zhang 2010; Berg 2014). Here we compare type and token frequencies, in order to evaluate these claims with fresh data.

1.3. Beats-and-Binding model of phonotactics

We investigate consonant clusters in the framework of the Beats-and-Binding phonotactic model established by Dziubalska-Kołaczyk (2002, 2009) which is embedded in Natural Linguistics (Dziubalska-Kołaczyk & Weckwerth 2002) and specifically in Natural Phonology. It is a syllable-less model, which explains the organization of consonant clusters in a language where beats constitute vowels (or the marked option of syllabic sonorants) and consonants are typically non-beats. A core of the Beats-and-Binding model is the Net Auditory Distance (NAD) Principle, which started as a modification of the Sonority Hierarchy principle (Whitney 1865; Sievers 1876; Jespersen 1904; Ohala 1990), called the Optimal Sonority Distance Principle (Dziubalska-Kołaczyk 2002: 82). The present NAD model offers the broadest existing possibility for defining degrees of intersegmental cohesion (Bertinetto et al. 2006) in terms of binding between the beat and adjacent non-beats and between adjacent non-beats, including the preferredness of a cluster.

NAD stands for the measure of auditory distances between neighbouring phonemes and allows construction of the hierarchy of preferences from the most to the least preferred cluster. A preference is understood as basically a universal preference which can be derived from more basic principles (Dressler 1999). A cluster is preferred if it satisfies a pattern of phonetic distances in terms of the place and manner of articulation plus the sonority between clusters specified by the universal preference relevant for their initial, medial or final position in the word (cf. Dziubalska-Kołaczyk 2009, 2014).

It is generally assumed that consonantal languages have more dispreferred consonant clusters than vocalic languages. In order to operationalize this assumption and to determine the status of consonant clusters in German, a software package, namely the Phonotactic Calculator developed by Dziubalska-Kołaczyk, Pietrala and Aperliński (2014) based on earlier work by Grzegorz Krynicki, can be applied. The default parameter values of the calculator include the manner of articulation (MOA), the place of articulation (POA), and a hierarchy of S/O (sonorant/obstruent) distinctions. Due to the Phonotactic Calculator's settings, the maximum number of consonant sequences to be analysed is bounded by triple clusters. Therefore, the current analysis of cluster preferredness in German is demonstrated based on triple consonant clusters.

Let us present the general predictions for a triple consonant cluster C1C2C3V, first for the word-initial position:

$\text{NAD (C1, C2)} < \text{NAD (C2, C3)} \geq \text{NAD (C3, V)}$

It reads: "For word-initial triple clusters, the NAD between the third consonant and the second consonant should be greater than or equal to the NAD between this third consonant and the vowel, and greater than the NAD between the second and the first consonant" (Dziubalska-Kołaczyk 2014: 5, also for the following citations).

For the word-final position VC1C2C3 it states:

$\text{NAD (V, C1)} \leq \text{NAD (C1, C2)} > \text{NAD (C2, C3)}$

The condition reads: "For word-final triple clusters, the NAD between the first consonant and the second consonant should be greater than or equal to the NAD between this first consonant and the beat, and greater than the NAD between the second and the third consonant."

The condition for medial triple clusters VC1C2C3V states:

$\text{VC1C2C3V NAD (V, C1)} \geq \text{NAD (C1, C2) \& NAD (C2, C3)} < \text{NAD (C3, V2)}$

It reads: "For word-medial triple clusters, the NAD between the first and the second consonant should be less than or equal to the NAD between the first consonant and the beat to which it is bound, whereas the NAD between the second and the third consonant should be less than between the third consonant and the beat to which it is bound."

The NAD product indicates a mean number of all the distances between the neighbouring phonemes in the cluster. It was introduced to the calculator in order to assign a preferability index which is "a number denoting a degree to which a given preference is observed" (Dziubalska-Kołaczyk 2019). The formula for word-initial consonant clusters is as follows:

NAD product = NAD C1C2 – NAD C2V

Thus, it allows the clusters to be ordered according to their degree of preferability values from the most preferred to the least.

1.4. Principles of Natural Morphology relevant for morphonotactics

Natural Morphology is a theory of preferences (Dressler 1999; Dressler & Kilani-Schoch 2017) divided into three subtheories. Of the first one, which accounts for universal preferences, the most relevant for morphonotactics are the parameters of iconicity (especially constructional diagrammaticity) and transparency. In connection with the subparameter of constructional diagrammaticity, German morphonotactic consonant clusters are nearly always due to affixation, which is the most iconic operation, whereas anti-iconic subtraction, as in *risk-ant* 'risky', derived from *Risiko* 'risk', is very rare (more in section 3). High transparency favours morphological decomposition, which is undertaken automatically in processing: also from this perspective, affixation facilitates decomposition more than word-internal modification and subtraction, and when a consonant cluster is only morphonotactic, the morpheme boundary is more salient, which facilitates decomposition or segmentation (cf. Korecky-Kröll et al. 2014). Also, high morphosemantic transparency facilitates decomposition, whereas opacity hinders it (Libben 1998; Gagné 2009: 264–268; Hongbo, Gagné & Spalding 2011; Dressler, Ketrez & Kilani-Schoch 2017). For example, the relationship between Ger. *Kun-st* 'art' and its verb base *könn-en* 'be able, can' is both morphotactically and morphosemantically obscure (cf. below and section 2.2).

Within the second subtheory, typological adequacy, German can be characterized as a weakly inflecting language, whose morphology is moderately rich (except in compounding). Thus, compounding may create more morphonotactic clusters than inflection or derivation. Unfortunately, we cannot investigate systematically word-internal clusters due to compounding because of our corpus; there is a lack of corpus linguistic tools for doing this semi-automatically. German is also a more suffixing than prefixing language. That inflectional prefixation cannot create consonantal clusters, corresponds to the type of suffixing language to which German belongs.

Within the third subtheory of system adequacy, the criterion of productivity (Bauer 2001; Dressler, Libben & Korecky-Kröll 2014) is very relevant: productive morphological rules, such as plural formation, inflection for person and past participle formation, are liable to be involved in many more morphonotactic consonant clusters than unproductive rules, such as deverbal action/result noun formation, such as in *Dien-st* 'service' and *Kun-st* (see above). The endpoint of non-productivity is reached in the case of fossil morphemes, such as the prefix in *Aber-glaube* 'superstition', where the base *Glaube* 'faith' is easy to detect. Still we can classify its internal triple consonant cluster /rgl/ as morphonotactic.

Although, from a semiotic point of view morphology is more important than phonology for morphonotactics (Dressler 1985, 1996a), diachronic change may transform morphonotactic clusters into phonotactic clusters, but not vice versa (cf. Dressler et al. 2019).

1.5. Database

The corpus linguistic research was based on the data extracted from the Austrian Media Corpus (AMC), which was developed at the Austrian Academy of Sciences (cf. Ransmayr, Mörth & Matej 2017). It is considered to be one of the largest corpus collections of the German language. It covers all printed resources from Austrian printed media for the last two decades, including the transcripts of Austrian television and broadcast news plus the news reports of the Austria Press Agency APA. This corpus contains about 40 million texts of various genres containing about 10 billion word tokens. It is linguistically annotated with morphosyntactic information and lemmatized. Due to its functionality, a list of all word types and word tokens containing the specific clusters in a given corpus can be selected along with the frequency of occurrence and part of speech. Clearly the numbers of types (inflectional word forms) given in the lists below refer to what is attested in the AMC; the number of potential correct forms is higher.

The starting point of the research was obtaining the data from the AMC. The corpus automatically allows identification of the position of a cluster, thus different queries were specified in the research. For instance, for the word-initial position the following query was involved "str.+". It reads word-initial triple cluster /str-/ followed by one or more character. Thus, all consonant clusters along with their frequency of occurrence in the corpus were retrieved, according to their position in the word, for further analysis. The next stage included the elimination of all irrelevant words, such as proper names, misspellings or non-words. The last stage of the analysis was the division of the words into three groups depending on whether the cluster is only morphonotactic, only phonotactic or both.

The second analysis related to measuring auditory distances in the cluster via the NAD calculator, which was introduced in the previous section. All examples are written in the national German orthography. In the German consonantal system, a phoneme <ch> is a voiceless palatal or velar fricative; <sch> (and word-initial <s> before a stop) is a voiceless sibilant. For the NAD calculator /r/ is specified as an uvular liquid approximant.

All clusters will be presented according to their position and each cluster will be exemplified by a single word, selected according to its high token frequency. If the number of word types occurred fewer than five times in the corpus, these words were eliminated from the analysis because most of them consisted of orthographic mistakes or they were non-words (especially names).

1.6. German phonotactics

The phonotactics of German consonant clusters has been described several times. Meinhold and Stock (1980: 180–188) include in their description differences between positions and observe the influence of morphology and of phonostylistics. Hirsch-Wierzbicka (1971) aims to present an exhaustive overview of consonant clusters, but limited to monosyllables. Thus, several word-initial and word-final triple and quadruple consonant clusters are missing (to some extent also for monosyllabic words). There are also incorrect statements about disallowed peripheral clusters. A classical generative account can be found in Heidolph, Flämig and Motsch (1981: 977–990) with the concept of the phonological structure conditions of morphemes (formatives) vs. words.

Szczepaniak (2010: 107) and Fehringer (2011: 97) found specific, but very limited corpus-based evidence that German seems to avoid long word-final morphonotactic consonant groups, insofar as a rising number of consonants correlates with a rising preference for the masculine and neuter genitive allomorph *-es* instead of the allomorph *-s*. This presupposes a continuum for cluster complexity, whereas Wiese (1988, 1991, 2000; cf. Orzechowska & Wiese 2011, 2015) makes a sharp distinction between marked extrametrical consonants (the third and fourth most peripheral consonant of a cluster) and the other consonants of a cluster (more in sections 2.5 and 4.2); loan words are considered to have more extrasyllabic consonants, i.e. more complex consonant clusters (cf. also section 3).

2. WORD-FINAL POSITION

In contrast to most Slavic and Romance and more conservative Indo-European languages, Germanic languages are rather rich in word-final consonant clusters, of both a phonotactic and a morphonotactic nature. Moreover word-final clusters are more complex and more numerous and more varied in types than word-initial ones.

The morphonotactic clusters occur in the final position in 2nd SG. person and are mainly represented by 3rd SG. verb forms, superlatives or past participles, as shown in Dressler and Dziubalska-Kołaczyk (2006; cf. Dressler et al. 2010). They end with the suffixes *-st* (2nd SG., superlative, plus the unproductive deverbal noun-forming suffix) and *-t* (3rd SG., past participle and denominal circumfixes derived from the past participle, ordinal-number-forming suffix).

2.1. Quadruple clusters

All word-final quadruple clusters consist of a sonorant and 3 obstruents, the two last being always /st/. All are either only morphonotactic or morphonotactic by default.

The following 20 clusters are only morphonotactic (always 2nd SG., sometimes also 3rd SG. or past participle):

/-lkst/ (5): *melk-st* '(you) milk', ver-folg-st '(you) persecute',

/-rkst/ (30): *merk-st* '(you) notice', *borg-st* '(you) borrow', past participle *ver-kork-st* 'messed up'. The only phonotactic case occurs in the noun *Gwirkst* that exists only in Austrian dialects and means 'tricky affair': this does not count for the standard.

/-mpst/ (11): *pump-st* '(you) pump', *plumps-t* '(s/he) flops' = *plumps-st* '(you) flop' (with obligatory degemination of /s+s/),

/-mp͡fst/ (10): *kämpf-st* '(you) fight',

/-nʃst/ (3): *wünsch-st* '(you) wish',

/-nt͡ʃst/ (3): *plantsch-st* '(you) splash', recent English loan words *launch-st*, *lunch-st.* In oral speech, the /s/ is most often reduced after /ʃ, t͡ʃ / when followed by /t/.

/-lfst/ (3): *hilf-st* '(you) help',

/-rfst/(65): *darf-st* '(you) may', *nerv-st* '(you) enervate',

/-rmst/ (29): *form-st* '(you) form'.

/-lmst/ (8): *film-st* '(you) film',

/-lxst/ (2): *strolch-st* '(you) roam about',

/-rxst/ (11): *schnarch-st* '(you) snore',

/-ft͜sst/ (2): *seufz-st* ‘(you) sigh’: normally the /s/ is fused with the preceding affricate,

/- xt͜sst (3): ächz-st ‘(you) groan’ (same fusion),

/-rt͜sst/ (2): *stürz-st* ‘(you) fall’ (same fusion),

/-l͜ʃst/ (2): *fälsch-st* ‘(you) falsify’, *feilsch-st* ‘(you) haggle’,

/-lt͜sst/ (1): *salz-st* ‘(you) salt’ (same fusion; 4 others potential, but not attested).

The following clusters are Gen.SG. of isolated masculine and neuter nouns:

/-ŋkst͜s/ (1*): Hengst-s* ‘stallion’ (masc.),

/-rpst͜s/ (1): *Herbst-s* ‘autumn’ (masc.), plus its numerous compounds,

/-lpst͜s/ (1): *Selbst-s* ‘the self’ (neuter),

/-rnst͜s/ (1): *Ernst-s* ‘earnestness’ (masc.), plus its numerous compounds.

The four following quadruple clusters are morphonotactic only as a strong default:

/-ŋkst/ as in *denk-st* ‘(you) think’ and in a variant pronunciation of *-ngst*, as in *sing-st* ‘(you) sing’, superlatives *jüng-st* ‘recently’, the morphosemantically somewhat opaque adverb *läng-st* ‘for a long time’ (closely related to the transparent superlative *der/die/das läng-st-e* ‘the longest’). However, there are two phonotactic exceptions: the nouns *Angst* ‘fear’ and *Hengst* ‘stallion’.

/-rpst/ occurs as a morphonotactic cluster in 2nd SG. verb forms in *stirb-st* ‘(you) die’, *wirb-st* ‘(you) advertise’ (and their preterits). The only phonotactic exception is *Herbst* ‘autumn’ and compounds thereof (with diachronic loss of a schwa, cognate with Engl. harvest).

/-lpst/ is only morphonotactic in *stülp-st* ‘(you) turn up (the collar)’ and *rülps-t* ‘(s)he burps’ = 2nd SG., Part. *ge-rülps-t*. The transitional exception is *selb-st* ‘oneself’ with a fossil suffix, related to *der/die/das-selb-e* ‘the same’.

/-rnst/ occurs as a morphonotactic cluster in 2nd SG forms, as in *lern-st* ‘(you) learn’, and as phonotactic only in the adj. *ernst* ‘earnest’ and its conversion into a noun.

Table 1 presents for each cluster the number of word types, its token frequency in the corpus and the type-token ratio. Since the NAD calculator is not able to measure all the distances within the quadruple clusters, no preferences can be deduced, but we chose the type-token ratio (TTR) calculation in order to arrive at some generalizations about the morphonotactic vs. phonotactic distribution of these clusters:

Table 1. Distribution of word-final quadruples

№	Cluster	Types	Tokens	TTR (%)
1	Vrpst	204	1,095,735	0.02
2	Vrfst	65	11,421	0.57
3	Vmpst	11	1,101	1
4	Vŋkst	37	9,688	0.38
5	Vrkst	30	5,149	11.38
6	Vrmst	29	255	10.38
7	Vrxst	11	106	1.29
8	Vmp͜fst	10	776	10.39
9	Vlmst	8	77	0.95
10	Vrnst	8	< 1,200,000	
11	Vlkst	5	526	0.82
12	Vnʃst	5	607	0.73
13	Vlpst	5	687	0.73
14	Vlfst	4	828	0.48
15	Vxt͜sst	3	3	100
16	Vlxst	2	2	100
17	Vlʃst	2	9	22.22
18	Vft͜sst	2	9	22.22
19	Vrt͜sst	2	4	50
20	Vnt͜ʃst	1	1	100
21	Vlt͜sst	1	1	100
22	Vŋkst͜s	1	23	4.35
23	Vrpst͜s	1	1,835	0.05
24	Vlpst͜s	1	25	4
25	Vrnst͜s	1	1,042	0.1

The type-token ratio is the most commonly used index of lexical diversity of a text, i.e. the number of tokens divided by the number of word types (McEnerny & Hardie 2012), which allows us to analyse the lexical variation of vocabulary containing a specific cluster in the corpus.

It can be observed that: 1) the overall number of tokens increases along with the number of word types); 2) the growth of tokens is exponential. Thus, relying on the data from the AMC corpus, it can be concluded that for word-final quadruple clusters the number of occurrences is in direct relation to the type frequency. Although there are also some other exceptions, there is a group of clusters /-lkst, -nʃst, -lpst/ which consist of a sonorant followed by an obstruent plus /st/. They are relatively rare in types, nevertheless they have a high token frequency in the corpus.

Based on the TTR, the groups of word-final quadruple clusters can be clearly distinguished according to 3 intervals: 1) 14 with a TTR between 0.02 and 1.29%; 2) 3 with a TTR between 10.38 and 11.38%; 3) for 4 clusters the TTR is exactly 100%. In addition, there are 2 with a TTR of 22.22%, 1 at 4.35% and 1 with a TTR of 50%. The TTR in /-rpst/ is the lowest, which means that there are very few words of very high frequency, e.g. *Herbst* 'autumn' is the most frequent word with the final cluster /-rpst/ in the corpus, the frequency of occurrences being due to a great number of compounds ending in *Herbst*. The second group consists of /-rkst, -rmst, -mp͜fst/, again due to the fact that there are rather few words that occur frequently. Finally, the TTR reaches 100% in the third group, where two words have just one form and two others two forms in the corpus. All clusters which are morphonotactic only as a strong default are in the first, the largest group.

The highest type and token frequency of /-rpst/ is due to the richness and productivity of German compounding which leads to the high occurrence of morphonotactic clusters in compounds with the final element *Herbst* 'autumn'. Thus, the TTR is by far the lowest of all the quadruple clusters. The next lowest TTR occurs in /-nkst/ which is the only quadruple cluster that includes a phonotactic cluster, i.e. in *Hengst* 'stallion' and its numerous compounds. Something similar to compounding takes place in productive particle word formation. But this pattern generates final verb clusters only in secondary clauses such as *Wenn du den Schal um-häng-st* 'if you put the scarf around (your neck)', and therefore the token frequency of such word-final morphonotactic clusters is very restricted and thus cannot compete with the number of phonotactic clusters in compounds.

Thus, the type-token ratio proves to be a far better distinguisher of quantitatively similar groups than the type or token frequency.

2.2. Triple clusters ending in *-t*

As expected, triple obstruent clusters are more numerous and varied than quadruple clusters. Not all of them, but nearly all start with a sonorant. In addition to the two final obstruents /st/ we also find /ft/ and combinations of all existing obstruents with final /s/, of course excluding prefinal /s/ due to degemination of /s+s/ and prefinal /d, t/ because of the fusion of the dental stop and /s/ to an affricate /t͜s/. Due to such fusion, genitives ending in /t͡s/ also exist, such as des *Punkt-s* 'of the point'. We exclude from our investigation triple clusters consisting of 2 sonorants and 1 obstruent, such as /-lmt, -lnt, -rnt/.

The exclusively morphonotactic triple clusters are 24 in number, i.e. 13 more clusters than the morphonotactic quadruple clusters:

/-xst/: *lach-st* '(you) laugh', superlative *höch-st* 'most highly',

/-xt͡st/: 3rd SG. ächz-*t* 'groans' and its participles,

/-fst/: *schaff-st* '(you) create', adverb *zu-tief-st* 'deepest', *nerv-st* '(you) get on nerves',

/-mst/: *träum-st* '(you) dream', *bums-t* '(s/he/you) bump(s)' and its participle, *spar-sam-st* 'most thriftily',

/-ʃst/: *wisch-st* '(you) wipe',

/-p͡fst/: *klopf-st* '(you) knock',

/-t͡ʃst/: *rutsch-st* '(you) slip',

/-f͡tst/: only in *seufz-t* '(s)he sighs' (and in the reduced 2nd person, see above, similarly in the following examples), and in the participle *ge-seufz-t*, and its derived verbs,

/-lft/: *hilf-t* 'helps', in weak past participles (e.g. *ge-golf-t* 'golfed'), and in *elf-t*, *zwölf-t* 'eleventh, twelfth',

/-lxt/: 3rd SG. and past participle *er-dolch-t* 'stabbed'

/- lt͡st/: *walz-t* '(s)he waltzes' and its participle,

/-nt͡st/: *tanz-t* '(s)he dances' and its participle, *ver-wanz-t* 'bug-infested', a circumfixation of *Wanze* 'bug',

/-lʃt/: only in *fälsch-t* '(s)he falsifies' and its participle and derived verbs,

/-mʃt/: only in *ramsch-t* '(s)he buys cheap junk' and its participle and derived verbs,

/-rt͡ʃt/ only in *turtsch-t* 'taps (eggs)' and its participle,

/-nʃt/: *wünsch-t* '(s)he wishes' and its participle,

/-pʃt/: *grapsch-t* 'grabs' and its past participle,

/-rʃt/: *forsch-t* '(s)he researches' and its participle,

/-nt͡ʃt/: *plantsch-t* '(s)he splashes' and its participle.

The following examples can never be the 2nd SG. (due to the phonological reduction of *-s*):

/-nxt/ in the only verb *tünch-t* 'whitewashes', its participles and its derivation into a particle verb,

/-lkt/: *melk-t* '(s)he milks', *folg-t* '(s)he follows' and their participles,

/-mp͡ft/: *kämpf-t* '(s)he fights' and its participle,

/-mpt/: *pump-t* '(s)he pumps', *bomb-t* '(s)he bombs' and their participles,

/-rpt/: *zirp-t* '(s)he chirps' and its participle, *stirb-t* '(s)he dies'

/-lpt/: *tülp-t* '(s)he turns up' and *wölb-t* 'curves' and their participles.

There are just 2 clusters which are morphonotactic as a strong default (if we take 75% of types as the criterion):

/-lst/: *will-st* '(you) want', *puls-t* '(s)he pulses' (and 2nd SG.) and its participle, adv. *schnell-st* 'most rapidly', but clearly phonotactic in Wulst 'bulge' and its compounds. Doubtful are *Schwul(-)st* 'bombast' and *Ge-schwul(-)st* 'tumour', because most people can relate it to the base verb *schwell-en* 'swell'. But this relation may be classified as rather metalinguistic; there is as yet no evidence that it would be active in processing (e.g. priming) experiments.

/-rt͡st/ as in *schmerz-t* 'it hurts' (also 2nd SG. *schmerz-st*) and its participle, but a unique phonotactic instance in *Arzt* 'physician' and its many compounds.

The following clusters are ambiguous with either a morphonotactic or a phonotactic majority:

/-nst/ as in *dien-st* '(you) serve' and in the homophonous noun *Dien-st* 'service' with an unproductive deverbal nominalization suffix, *grins-t* '(s) he grins' (plus 2nd SG.) and its participle, adv. *fein-st* 'in the finest way'. The cluster is clearly phonotactic in *ernst* 'earnest', *sonst* 'otherwise', *Wanst* 'paunch'. We should also add earlier derivations such as *Kunst* 'art' which many relate metalinguistically, against furious artist's opposition, to the verb *könn-en* 'to be able'; *Gunst* 'favour', which few relate metalinguistically to the etymologically cognate verb *gönn-en* 'not begrudge smth to smbd'; similarly *Brunst* 'sexual heat' to *brenn-en* 'burn'. In terms of types (excluding compounds), the cluster /-nst/ might be called morphonotactic by default, but the 1,993 compounds with the second element *-kunst* render the global type and token frequency of phonotactic clusters the majority.

/-rst/ is morphonotactic in cases such as *war-st* '(you) were', the superlative adverb *schwer-st* 'heaviest', isolated *mors-t* '(s/he/you) send in Morse' and its participle vs. phonotactic *Wurst* 'sausage', *Forst* 'forest', *Durst* 'thirst', *erst* 'first' (which, like its English correspondent, was originally a superlative), but most types occur in compounds. *Ober(-)st* 'colonel' is thoroughly lexicalized (morphosemantically opaque), but clearly related to the superlative *der ober-ste* 'the highest'. When excluding compounds, the types are morphonotactic by default.

/-pst/ is morphonotactic in cases such as *tipp-st* '(you) type', *lieb-st* '(you) love', *pieps-t* '(s)he peeps' (also 2nd SG. and particple *ge-pieps-t*), superlative (or, more precisely, excessive) adverb *herz+aller-lieb-st* 'wholeheartedly dearest', phonotactic in *Papst* 'pope', *Obst* 'fruits', *Probst* 'provost'. Again, this cluster can be considered to be morphonotactic by default, when excluding compounds, but the abundant metaphoric compounds of *Papst* make the global type frequency and token frequency of phonotactic clusters majoritarian.

/-rkt/ occurs as a morphonotactic cluster in *merk-t* '(s)he notices' *sorg-t* '(s)he cares' and their participles, but as a phonotactic cluster in *Markt* 'market', *Infarkt* 'infarct' and their numerous compounds. Without these the cluster is morphonotactic by default.

/-ŋkt/ (written with also *-ngt)* is morphonotactic by default as in *bring-t* '(s)he brings', if one excludes the noun *Punkt* 'point, dot' with its numerous compounds, again as the richness of German compounding type and token frequency hides the basic default. Another noun with the phonotactic cluster is *Instinkt*.

/-rxt/ (phonetically [rçt]) is similarly morphonotactic by default, as in *ge-pferch-t* 'crammed', with the only phonotactic cluster in *Furcht* 'fear' and its numerous compounds.

/-rft/ is similarly morphonotactic by default, as in *wirf-t* 'throws' and *nerv-t* 'enervates', with the phonotactic exceptions *Werft* 'wharf' with its many compounds and *Notdurft* 'need' (where the earlier morpheme boundary before nominalizing *t* is obsolete).

/-nft/ is the only cluster of this subgroup which is phonotactic by default, as in *sanft* 'mild' (Austrian variant *Senft* 'mustard' with a secondarily attached final /t/). The only morphonotactic exception is the ordinal number *fünf-t* 'fifth', whereas it is improbable that an analogous morpheme boundary is processed in *Brunft* 'rut (of deer)', historically derived from *brenn-en* 'to burn', because of its morphotactic and morphosemantic opacity, and with most nouns analogously derived from particle verbs with the verbal base *komm-en* 'come', such as *Zukunft, Hinkunft* 'future' vs. *zukommen* 'approach, belong'.

/-kst/ (also written -*chst*, -*ckst*, -*gst, -xt*) is morphonotactic by default, as in *wächs-t* 'grows' (also in the 2nd singular *weck-st* '(s)he awakes'), the only phonotactic exceptions are *Text* 'text' and *Axt* 'axe' with their numerous compounds.

There are no other word-final triple consonant clusters with 2 final obstruents, unless in foreign names, such as *Minsk, Kursk*. Other comparable triple clusters with final *-t* do not occur, because conceivable and pronounceable clusters such as *-skt, -spt* do not occur as phonotactic clusters and, in contrast to English, they are excluded as morphonotactic clusters, because no verb roots (nor nouns) ending in *-sk, -sp* exist in German. Adjectives ending in *-sk* do not form a superlative in *-sk+st*, but insert an *-e-* before the superlative suffix. Other fricatives have a still smaller phonotactic distribution than /s/.

Thus, all word-final triple clusters, which contain two obstruents are morphonotactic (only exception: those in *-nft*), because phonotactic clusters either do not occur or only occur as the exceptions when counted in lemmas. But their type and token number may be competitive with morphonotactic ones due to compounding. Many of the lemmas with final phonotactic clusters go back to derivations with a morphonotactic cluster.

As expected, morphonotactic clusters ending in the longer suffix *-st* have fewer phonotactic counterparts than morphonotactic clusters ending in the shorter suffix *-t*.

Turning to a NAD analysis of triple final clusters ending in /t/, we start with the presentation of the frequency demonstrated in Table 2:

Table 2. Frequency ranks of word-final triples

№	Cluster	Types	Tokens	TTR (%)
1	ŋkt	6,196	9,831,812	0.063
2	nst	5,594	5,487,640	0.1
3	kst	2,136	2,457,398	0.09
4	nft	1,640	2,601,645	0.06
5	rst	1,401	5,649,995	0.02
6	rt͜st	1,226	1,399,699	0.09
7	pst	845	4,776,987	0.02
8	lst	360	92,894	0.4
9	rft	304	597,052	0.05
10	nt͜st	266	560,076	0.05
11	xst	246	1,838,731	0.01
12	mp͜ft	232	662,652	0.03
13	mst	226	164,703	0.14
14	lkt	182	2,809,304	0.01
15	rʃt	163	625,920	0.03
16	lt͜st	156	54,562	0.29
17	fst	136	89,308	0.15
18	rkt	134	1,358,674	0.01
19	rxt	104	87,843	0.12

20	mpt	98	157,983	0.06
21	rpt	90	294,409	0.03
22	lpt	50	11,632	0.43
23	lft	46	376,380	0.01
24	nʃt	45	354,583	0.01
25	ʃst	40	1,388	2.89
26	t͜ʃst	31	150	20.7
27	pʃt	27	5075	0.41
28	ft͜st	23	19,692	0.12
29	lʃt	22	52,109	0.04
30	xt͜st	19	9,353	0.2
31	nt͜ʃt	19	9,188	0.21
32	lxt	16	1,580	1.01
33	mʃt	10	770	1.3
34	nxt	5	2066	0.24
35	rt͜ʃt	4	4	100
36	p͜fst	16	374	4.28

In contrast to quadruple clusters, triple clusters do not form several neatly separated groups according to the TTR: the TTR of just 4 clusters is clearly above 1%, one amounts to 20.7% and only one has a TTR of 100%. None of the triple clusters hast just 1 type.

The NAD phonotactic calculator establishes the preferences of the clusters (structure VCCC) as presented in Table 3:

Table 3. Preference rankings of word-final triples according to NAD[3]

№	IPA transcription	NAD (VC)	NAD (C1C2)	NAD (C2C3)	NAD product	Preferred cluster?
1	Vrpt	2	6.6	1	5.1	Yes
2	Vrt͜st	2	5.1	0.5	3.85	Yes
3	Vrft	2	5.1	1.5	3.35	Yes
4	Vrst	2	4.6	1	3.1	Yes

[3] Three clusters /-nt͜ʃt/, /-t͜ʃst/ and /rt͜ʃt/ were excluded from the analysis because the NAD calculator does not recognize affricate /-t͜ʃ/. Therefore, they were counted manually.

5	Vlkt	2.5	4.8	1.3	2.9	Yes
6	Vlpt	2.5	4.5	1	2.75	Yes
7	Vrkt	2	4.3	1.3	2.65	Yes
8	Vrʃt	2	4.1	1.5	2.35	Yes
9	Vlxt	2.5	5.5	4	2.25	Yes
10	Vnkt	3	4.3	1.3	2.15	Yes
11	Vnxt	3	5	4	1.5	Yes
12	Vlt͜st	2.5	3	0.5	1.5	Yes
13	Vrxt	2	4.4	4	1.4	Yes
14	Vmʃt	3	3.5	1.5	1.25	Yes
15	Vmp͜ft	3	3	1	1	Yes
16	Vlft	2.5	3	1.5	1	Yes
17	Vmst	3	3	1	1	Yes
18	Vmpt	3	3	1	1	Yes
19	Vlʃt	2.5	3	1.5	1	Yes
20	Vxt͜st	5	3.5	0.5	0.75	No
21	Vnt͜st	3	2.5	0.5	0.75	No
22	Vlst	2.5	2.5	1	0.75	Yes
23	Vnft	3	2.5	1.5	0.25	No
24	Vnʃt	3	2.5	1.5	0.25	No
25	Vxst	5	3	1	0	No
26	Vnst	3	2	1	0	No
27	Vkst	6	2.3	1	-1.2	No
28	Vpʃt	6	2.5	1.5	-1.25	No
29	Vpst	6	2	1	-1.5	No
30	Vft͜st	5	1	0.5	-1.75	No
31	Vp͜fst	5.5	1	1	-2.25	No
32	Vfst	5	0.5	1	-2.5	No
33	Vʃst	5	0.5	1	-2.5	No

From Table 3 the following conclusions can be drawn:

The majority of preferred clusters start with a rhotic, lateral or nasal sonorant followed by two obstruents or another sonorant. The most significant distance between the neighbouring phonemes is always greatest when it starts with a rhotic or lateral sonorant, for instance the NAD product of /rpt/ is 5.1 and the NAD product of /rt͜st/ is 3.85.

Out of 33 word-final consonant clusters, 19 clusters are preferred and 14 dispreferred. If we add the 3 clusters that the NAD calculator could not handle, then we obtain 19 preferred clusters and 17 dispreferred clusters.

However, there is the question of whether similar predictions can be deduced in a simpler process of calculation. Since the NAD calculator is the most elaborate tool for deducing the predictions on the degrees of markedness for (mor)phonotactic clusters so far, it is worth trying to modify the method of NAD calculation.

Thus, we applied a factor analysis in order to test whether there is a correlation among the variables which were previously obtained in the present research. For the factor analysis, 30 word-final consonant clusters were selected and 7 independent variables. The first and second variables are the number of the word types and tokens from the AMC for each cluster followed by the auditory distances between the neighbouring phonemes according to the NAD calculator. The next two variables represent the information whether the cluster is preferred or dispreferred and the division between phonotactic vs. morphonotactic (Phon/morph) clusters as presented in Table 4.

Table 4. Factor analysis for word final triple consonant cluster

	Factor loadings (Varimax normalized) Extraction : Principal components (Marked loadings are > .700000)		
Variables	Factor (1)	Factor (2)	Factor (3)
Types	-0.024865	**-0.924373**	0.073853
Tokens	-0.142575	**-0.849302**	0.216074
NAD (VC)	**-0.832629**	-0.042204	0.403198
NAD (C1C2)	**0.916486**	-0.114682	-0.090996
NAD (C2C3)	0.051592	0.071495	**-0.966928**
Preferences	**0.918758**	0.024393	0.194894
Phon/morph	-0.283889	**0.730136**	0.201906
Expl. var	2.481534	2.129518	1.236691
Prop. of total. var	0.354505	0.304217	0.176670

Numbers in bold indicate a significant correlation among the variables. For instance, in Factor (1) we may observe a certain correlation between NAD (VC) and NAD (C1C2). The possible explanation is that if we look at the NAD table of all 30 clusters, we can see that the measures of NAD (VC) and (C1C2) are inversely proportional to each other in most of the cases. For instance, if the NAD (VC) is high then the NAD (C1C2) will be smaller. For example, in the word-final cluster Vfst the NAD (VC) is equal to 5 and the NAD (C1C2) is 0.5. And conversely, if we take the cluster Vrpt, where the NAD (C1C2) is equal to 6.6 and NAD (VC)=2.

The next observation is that cluster preferredness is related to the NAD (VC) and the NAD (C1C2). In general, if the NAD (C1C2) is higher than the NAD (VC), then the cluster is more likely to be preferred. This corresponds entirely to the NAD formula for triple finals shown above.

From Factor (2) we can see that there is a certain correlation between word types and tokens. They are connected in the same direction, so we could assume that if the number of word types grows, then the frequency grows as well.

For Factor (3) we can observe that the NAD (C2C3) is not connected to any of the variables, but it is still significant, presumably to other variables not yet discussed.

Most notably, the factor analysis has shown that the NAD (C2C3) is not related to the NAD (VC) or the NAD (C1C2), which goes against a well-established NAD formula for predicting the preferredness for word-final triple clusters. Therefore, one assumption that can be inferred is that the NAD distances of two phonemes in the cluster, namely the NAD (VC) and the NAD (C1C2) might be enough to decide on the preferredness of word-final clusters in German. However, more research on consonant clusters in different word positions as well as of different languages is needed in order to corroborate this statement. For that reason, we have compared the cluster preferredness of German, English and Polish in the word-initial and word-final positions via the NAD calculator when the most peripheral consonants were excluded from the analysis. The results are discussed in section 4.2.

If we compare the preference predictions in Table 3 or just compare its third and fourth columns, where the NAD (C1C2) should be bigger than the NAD (VC), and if we split Table 2 into two based on the frequency ranking, putting 18 clusters into the first half and 18 into the second, then we find 11 preferred and 7 dispreferred clusters within the first group, and 10 preferred and 8 dispreferred clusters in the second half. This is a positive, i.e. supportive, but not a significant difference. With regard to the claim that phonotactic clusters are more preferred than morphonotactic clusters, we found that among the exclusively morphonotactic clusters, 14 are preferred and 11 dispreferred, whereas among those clusters which are both morphonotactic and phonotactic, 7 are preferred and 4 dispreferred. This is again a positive but not a significant difference.

Moreover, all (but one) of the word-initial triple clusters, which are all exclusively phonotactic, are preferred clusters. And this seems to represent a very significant difference from the mainly morphonotactic word-final clusters. However, the triple final clusters ending in *-s* (discussed in the following section 2.3) are all exclusively morphonotactic and all preferred clusters.

2.3. Triple clusters ending in *-s*

A further source of word-final morphonotactic obstruent groups is the nominal -s Gen.SG., less commonly the homophonous plural suffix as in *Kalb-s* 'calf', (also plural), *Korb-s* 'basket', *Ge-zirp-s* 'chirping', *Schilf-s* 'reed', *Dorf-s* 'village', *Nerv-s* 'nerve', *Talg-s* 'tallow'. Parallel phonotactic clusters occur in *Rülps* 'belch' and *Mumps*. Similar morphonotactic clusters arise through the suffixation of plural *-s*, as in Gen.SG. and PL *Tank-s*, *Skalp-s* 'scalp', *Ulk-s* 'trick', and adverbial *-s,* as in *aller-ding-s* 'indeed'.

Word-final, exclusively morphonotactic, triple clusters with /s/ at the end are the following (all Gen.SG., if also plurals, then explicitly noted):

/-rps/: *Bewerb-s* 'competition', *Korb-s* 'basket' and their numerous compounds,

/-rfs/: *Dorf-s* 'village', *Wurf-s* 'throwing' and *Nerv-s* 'nerve' and their numerous compounds,

/-rks/ as in Gen.SG. *Bezirk-s* 'district', Gen.SG. and PL of recent English loan-words, such as *Park-s*. A phonotactic exception is *Murks* 'botch',

/-rxs/: *Monarch-s* with a few compounds,

/-rʃs : *Hirsch-s* 'stag',

/-lfs/: *Wolf-s* 'wolf',

/-lks/: *Erfolg-s* 'success', *Volk-s* 'people, folk',

/-lxs/: *Elch-s* 'elk' with several compounds,

/-nks/: also PL in the English loan word *Song-s*, only adverb *link-s* 'to the left,'

/-nʃs/: *Wunsch-s* 'wish' with a few compounds,

/-nxs/: only *Mönch-s* 'monk' with its many compounds.

/-nt͜ʃs/: only in English loan words, e.g. *Brunch-s* (more than 60% plurals, less than 40% Gen.SG. in the average),

/-mps/: only in English loan words (also PL), e.g. *Vamp-s*; a phonotactic exception is the loan word *Mumps*,

/-lps/ occurs only in *Kalb-s* 'calf' and in the loan word (also PL)

Skalp-s ‘scalp’ and their compounds; a phonotactic exception is the onomatopoeic *Rülps* ‘belch’,

/-mp͡fs/: *Kampf-s* ‘fight’ and its compounds,

/-mʃs/ only in *Ramsch-s* ‘junk’,

/-sks/ only in loan words (also PL), e.g. *Disk-s*.

The frequency ranking of these clusters is presented in Table 5:

Table 5. Frequency ranks of triple clusters ending in *-s*

№	Cluster	Types	Tokens	TTR
1	ŋks	10,218	5,608,107	0.18%
2	rks	4,398	858,787	0.51%
3	rfs	1,175	189,687	0.62%
4	rps	1,165	94,392	1.23%
5	lks	506	76,976	0.66%
6	lfs	56	13,961	0.4%
7	mp͡fs	20	35,000	0.56%
8	rxs	7	70,000	0.01%
9	rʃs	7	3,300	0.21%
10	lxs	6	268	2.24%
11	nt͡ʃs	5	370	1.35%
12	sks	5	145	3.45%
13	nʃs	2	176	1.14%
14	mʃs	1	6	16.7%

The spread of the TTR is similar to the triple clusters ending in /t/, but there is one cluster with only one type.

The preferences established by the NAD calculator for VCCC clusters are the following (see Table 6):

Table 6. Preference rankings of word-final triples ending on *-s* according to NAD

IPA transcription	NAD (VC)	NAD (C1C2)	NAD (C2C3)	NAD product	Preferred cluster?
Vrps	2	6.6	2	4.6	Yes
Vrfs	2	5.1	0.5	3.85	Yes
Vlks	2.5	4.8	2.3	2.4	Yes
Vlps	2.5	4.5	2	2.25	Yes

Vrks	2	4.3	2.3	2.15	Yes
Vlfs	2.5	3	0.5	1.5	Yes
Vŋks	3	3	2.3	0.35	Yes

Thus, all triple clusters ending in *-s* are preferred clusters, although all of them are exclusively morphonotactic, two of them with a marginal phonotactic exception.

Also, there are several morphonotactic double final morphonotactic consonant clusters with an affricate /t͜s/, due to Gen.SG. and rarely PL *-s*: /xt͜s/ as in *Berichts* 'report', /kt͜s/ as in *Projekts* 'project', /pt͜s/ as in *Konzepts* 'concept', /lt͜s/ as in *Anwalts* 'lawyer', /nt͜s/ as in *Abends* 'in the evening', and /rt͜s/ as in *Jahrhunderts* 'century'. The only phonotactic correspondents are words such as *Holz* 'wood', *Tanz* 'dance', *Scherz* 'joke', i.e. if a sonorant precedes an affricate.

A problem is represented by imperatives of the type *knicks*! 'curtsey!', *schubs*! 'push!'. First, it is unclear whether the word-final *-s* is synchronically still a derivational suffix. Second, even if not, it is unclear whether such imperatives are to be classified as base forms (if yes, then phonotactic) or as morphologically derived from the infinitive as a lexical entry.

2.4. Triple clusters ending in -t͜s

The masculine and neuter Gen.SG, *-s* (potentially, also of the homophonous plural suffix, but actually only in a single cluster) is the source of nearly always morphonotactic clusters ending in the affricate *-ts* due to fusion of the inflectional suffix with a stem-final dental stop (for frequency ranks see Table 7):

/-rst͜s/: *Durst-s* 'thirst',
/-lst͜s/: *Schwulst-s* 'bombast',
/-pst͜s/: *Papst-s* 'pope', *Herbst-s* 'autumn' and their many compounds,
/-nst͜s/: *Dienst-s* 'service' and its many compounds,
/-rkt͜s/: *Markt-s* 'market' and its many compounds,
/-nkt͜s/: *Punkt-s* 'point' and its many compounds,
/-nft͜s/: *Senft-s* 'mustard',
/-rpt͜s/ only in *Exzerpt-s* 'excerpt',
/-tst͜s/ only in *Arzt-s* 'physician' with its many compounds,
/-kst͜s/ only in *Text-s* and its compounds

Table 7. Frequency ranks of triple clusters ending in *-ts*

№	Cluster	Types	Tokens	TTR
1	rst͜s	8	2,434	0.33%
2	psts	5	3,085	0.16%
3	lst͜s	5	389	1.29%
4	nst͜s	4	8,800	0.04%
5	rkt͜s	2	58,097	0.003%
6	nfs	2	301	0.66%
7	kst͜s	1	2,000	0.05%
8	rtst͜s	1	407	0.24%
9	rpt͜s	1	70	1.43%

Here we have no groupings of clusters according to TTR, but there are three clusters with just one type. Again, all clusters are preferred according to the NAD calculator, although all of them are exclusively morphonotactic.

2.5. Word-initial position

The German standard has no monoconsonantal prefixes, in contrast to Bavarian-Austrian dialects, as in *g'storben* 'died', *b'soffe*n 'drunk', *z'ruck* 'back(wards)' etc., corresponding to Standard German *ge-storb-en*, *be-soff-en*, *zu(-)rück*. Thus, the German standard is rather poor in word-initial clusters, all word-initial clusters are exclusively phonotactic. Some of the more dispreferred ones occur only in loan words from Ancient Greek and their derivations, e.g. /mn-/. German phonotactic initial double clusters were partially studied in Dziubalska-Kołaczyk (2002) with regard to universal phonotactic preferences. Moreover, double obstruent clusters serve as a basis for the complexity of triple initial clusters.

Phonotactic preferences for word-initial clusters in German have been studied by Orzechowska and Wiese (2011, 2015). They proposed an alternative approach to the NAD which is not limited to the size of the cluster and is not based on a sonority hierarchy but on an empirical analysis of features. The analysis of German initial clusters was based on 15 parameters, which included different values such as the cluster complexity, place of articulation, manner of articulation and voicing, in order to build a quantitative ranking of all clusters in terms of adherence to the preferences established by the Sonority Sequencing Generalization. This last approach will not be followed here.

For our study, the most interesting word-initial double clusters consist of two obstruents, particularly with a fricative in first position and a stop in second position: /ʃt-/ as in *statt* ‘instead of’ and /ʃp-/ as in *spielen* ‘to play’. Words of foreign origin can also start with /sk-/ as in *skeptisch* ‘sceptical’, /sp-/ as in *Spatium* ‘space’, /st͡s-/as in *szenisch* ‘scenic’, isolated /xt-/ as in *chthonisch* ‘chthonic’, and /ft-/ as in *Phthisis* ‘wastage’.

A fricative is followed by another fricative, or rather approximant, in /ʃv-/ as in *schwer* ‘heavy’, or in loan words in /sv-/ as in *Sweater*, /sf-/ as in *sphärisch* ‘spherical’, or /sx-/ as in *Schizophrenie* ‘schizophrenia’, and by an affricate in /t͡sv-/ as in *zwei* ‘two’.

An obstruent is followed by a sonorant, first as a fricative, as in *schreiben* ‘to write’, /ʃm-/ as in *schmecken* ‘to taste’, /ʃn-/ as in *schneiden* ‘to cut’, /ʃl-/ as in schließen ‘to close’, /fl-/ as in *flach* ‘flat’, /fr-/ as in *fragen* ‘to ask’, /vr-/ as in *Wrack* ‘wreck’, only in loan words /sm-/ as in *Smaragd* ‘emerald’, /xr-/ only in the isolated learned loan word *Chrie* ‘school theme’, (/vl-/ only in foreign names such as *Vladimir, Wladiwostok*).

A stop is followed by a sonorant in /gr-/ as in *groß* ‘large’, /gl-/ as in *glücklich* ‘happy’, /gn-/ as in *gnadenlos* ‘merciless’, /kl-/ as in *Kleid* ‘dress’, /kr-/ *krank* ‘sick’, /kn-/ as in *Knie* ‘knee’, /bl-/ as in *bleiben* ‘to stay’, /br-/ as in *brechen* ‘to break’, /pl-/ as in *plump* ‘clumsy’, /pr-/ as in *Pracht* ‘splendour’, /dr-/ as in *drei* ‘three’, /tr-/ as in *tragen* ‘to wear’. An affricate is the first obstruent in p͡fl as in *pflegen* ‘to care for’, /p͡fr-/ as in *pfropfen* ‘to graft’.

A stop is followed by a fricative in words of foreign origin in /ks-/ as in *Xenophobie* ‘xenophobia’ or /ps-/ as in *psychisch* ‘psychological’. A stop is followed by the fricative or approximant /v/ in /kv-/as in *Quelle* ‘source’, or by an affricate in /t͡sv-/ as in *Zwang* ‘coercion’.

A sequence of word-initial stops is limited to words of Ancient Greek origin: /pt-/ as in *Pteridin* ‘pteridine’, /kt-/ as in *ktenoid* ‘ctenoid’.

The majority of double clusters that do not occur only in learned words of foreign origin respect the preferences of the Beats-and-Binding-Model (Dziubalska-Kołaczyk 2002: 112).

In this contribution, we stick to the longer clusters with the maximum number of consonants in the onset, which is three. There are eight types of triple initial consonant clusters in German (see Table 8). All of them consist of two obstruents plus a sonorant or approximant: /ʃtr-/ as in *streng* ‘strict’, /ʃpr-/ as in *spricht* ‘s/he speaks’, /ʃpl-/ as in *Splitter* ‘splinter’; next in words of foreign origin /skr-/ as in *skrupellos* ‘ruthless’, /skl-/ as in *sklavisch*, adjective of ‘slave’. In more recent loan words we find also /skv-/ as in *Squaw* (the only integrated loan word with this cluster, with the possible exception of *squash*), /spr-/ as in *Sprinter* and /spl-/ as in *Spleen*.

Table 8. Frequency ranks of triple word-initial clusters

№	Cluster	Types	Tokens	TTR (%)
1	ʃtr	15,371	2,451,048	0.63
2	ʃpr	6,317	2,861,933	0.22
3	skr	782	26,878	3
4	skl	221	3,175	7
5	ʃpl	104	6,131	1.7
6	spl	97	6,013	1.61
7	spr	25	15,420	0.16
8	skv	1	1,845	0.05

These triple clusters also exhibit no grouping according to TTR; only one cluster has just one type.

Table 9. Preference rankings of word-initial triples according to NAD

IPA transcription	NAD (C1C2)	NAD (C2C3)	NAD (CV)	NAD product	Preferred cluster?
sprV	2	6.6	2	4.60	Yes
ʃprV	2.5	6.6	2	4.35	Yes
ʃtrV	1.5	5.6	2	3.85	Yes
sklV	2.3	4.8	2.5	2.4	Yes
splV	2	4.5	2.5	2.25	Yes
skrV	2.3	4.3	2	2.15	Yes
ʃplV	2.5	4.5	2.5	2	Yes
skvV	2.3	2.8	5	-0.85	No

Table 9 presents the NAD analysis of these clusters and the quantification of rising preferences. For word-initial consonant clusters we undertook an analogous factor analysis as for the word-final consonant clusters in section 2.2. When eliminating the first consonant, the two remaining NAD distances, NAD (C2C3) and NAD (CV), again showed the same preferences as when including the first consonant, i.e. we arrived at the same result as in section 2.2.

In conclusion we can see that:

1) All word-initial triple clusters consist of initial double obstruent clusters of a s(h)ibilant plus a stop followed by a rhotic or lateral sonorant or the fricative/approximant /v/. Other double clusters which occur in the

word-initial position, i.e. /bl, br, gr, gl, gn, gm, dr, xr, xt, kn, p͡fl, p͡fr, ʃl ʃv, ʃr, ʃm, ʃn, ps, sf, sm, st͡s, t͡sw/ cannot be part of a word-initial triple cluster, except for extragrammatic words such as the interjection *pst*, which has the further irregularity of containing a syllabic fricative.

2) There is a moderate correlation between the degree of preferredness and the frequency in the AMC: the most preferred cluster is /ʃpr/, which has the highest token frequency and the second-highest type frequency; the next cluster in the hierarchy of preferences is /ʃtr/, which has the highest type frequency and the second-highest token frequency. The other three clusters differ little in preferredness and their frequency ranks decrease in parallel for types and tokens. The reason for the mismatch between the type and token frequency differences of /ʃtr/ and /ʃpr/ is on the one hand historical, insofar as they go back to the earlier clusters /str/ and /spr/, the only word-initial triple consonant clusters reconstructed with some certainty for Proto-Indo-European (Oppermann 2004). On the other hand, the general phonotactic preference for /ʃpr/ may have had a positive impact on its token frequency. The only dispreferred cluster /skv/ is rare and occurs only in one word type (or two).

3. WORD-INTERNAL POSITION

Word-internal clusters are presented only briefly and selectively for the following reasons: first of all, word-medial consonant clusters are much more varied and complex than initial and final ones, so that an equally extensive study would exceed space limits. Second, the corpus linguistic tools of the AMC do not permit the same procedures of analysis as for initial and final clusters. Third, the NAD calculator cannot predict preferences for the many complex clusters of more than three consonants. Fourth, internal clusters are psycholinguistically less important than peripheral clusters due to the bathtub effect, which renders the periphery of a unit better perceivable than its interior (Aitchison 2003: 138). Therefore, we limit our discussion to observations of general differences between morphonotactic and phonotactic consonant clusters and their explanations.

It holds for phonotactic clusters that word-internal syllable onsets always follow the pattern of word-initial onsets. In compounding and derivation, the syllable boundary always follows the morpheme boundary in consonant clusters.

In a word-internal position, there is a much greater variety of consonant clusters than in the peripheral positions. Phonotactic clusters that occur only word-internally have an internal syllable boundary, but they are rather few, such as /fk, dl, dv/ as in the plant name *Levkoje*, in *Adler* 'eagle', where a vowel has been lost, and *Advent* 'advent', where a morpheme boundary has been lost, and /tl/ as in the loan word *Atlas*. There are a few triconsonantal phonotactic clusters, such as /ktr, ltr, mpl, rt͡sn, stm, / as in the loan words *Spektrum, Altruismus* 'altruism', *Amplitude* 'amplitude', *Arznei* 'medicine', *Asthma*, thus hardly any with two obstruents.

The bulk of new word-internal consonant clusters are morphonotactic due to the addition of morpheme-initial to morpheme-final clusters in compounding and affixation. This often creates morphonotactic clusters which are disallowed word-initially or word-finally and may contain more consonants than are permitted in the word periphery. Examples are the compound *Herbst+pflanze* 'autumn plant' and the suffixation *herbst-lich* 'autumnal', as well as the prefixation *ent-springen* 'originate'. In compounding, interfixation may either break up (by the interfix -*e*-) or increase (by the much more frequent interfix -*s*-) the sequence of consonants as in *Weg+e+lagerer* 'highwayman' and *König+s+schloss* 'royal castle'. The syllable boundary is always after the interfix, which fits with the fact that the main morpheme boundary is always after, and never before, the interfix.

Verb prefixation and particle verb formation creates new word-internal consonant clusters as well. For example, the separable particle *ab*- motivates the exclusively morphonotactic clusters /p-d, p-t, p-g, p-k, p-ʃ, p-t͡s, p-v/, as in *ab-drehen* 'turn off', *ab-geben* 'give in', *ab-kommen* 'get away', *ab-treten* 'wear out', *ab-schaffen* 'abolish', *ab-wickeln* 'unwind', *ab-ziehen* 'remove', (with the addition of longer clusters, as in *ab-streiten* 'deny'). Moreover, some of the few non-separable verbal prefixes create new clusters, as with *ent*-, and the earlier but now only vestigial affix *ant*- as in *Ant-wort* 'answer'; in the parallel formation *Antlitz* 'face' the morpheme boundary was lost, and the cluster became a phonotactic one. A morpheme boundary must also be assumed after cranberry morphs, as in *Sint-flut* 'deluge', cf. *Flut* 'flood'.

In contrast to many non-Germanic Indo-European languages, German affixation does not provoke internal vowel deletion and internal morphonotactic clusters caused by it, other than of the weakest vowel schwa. An exception is *Risiko* 'risk' → adj. *risk-ant*. An epenthetic schwa is lost before a (originally word-final) sonorant in derivation, such as in the derived adjectives *adl-ig* 'noble', *silbr-ig* 'silvery' (more examples in Meinhold & Stock 1980: 197–201). Inflectional affixation results even more rarely in subtraction, which creates morphonotactic clusters, such as in *Risk-en*, the plural of *Risiko* (in contrast to the much greater frequency in Slavic languages, Latin, Greek and other ancient Indo-European languages).

In addition, word formation creates geminate consonants which are disallowed morpheme-internally, and phonotactically, with even more marked results; pseudogeminates are created by syllable- and morpheme-final obstruent devoicing, as in *ab-bauen* 'dismantle' with /p, b/.

Among clusters which are both phonotactic and morphonotactic, the productive word formation devices of compounding, verbal prefixation and particle verb formation may greatly outweigh the proportion of phonotactic clusters in types and tokens, e.g. for clusters starting with /-st-/, as in *west+römisch* 'Western Roman' and *aus-treiben* 'drive out' as opposed to phonotactic cases in loan words, such as *Pastrami*. This may create problems for matching phonotactic and morphonotactic clusters in psycholinguistic tests.

Only the complexity of consonant clusters, at least in terms of the number of member consonants and of the creation of new clusters which are not allowed in phonotactics, rises due to morphological operations. And in this sense, morphonotactic clusters are, on average, more marked than phonotactic clusters.

4. CONCLUSIONS

4.1. General results

The claim that in general morphonotactic clusters are more dispreferred than phonotactic clusters (Dressler & Dziubalska-Kołaczyk 2006: 83, Zydorowicz et al. 2016: 19–20) has been disproven for German peripheral triple consonant clusters. This removes an apparent contradiction between the claim and external psycholinguistic evidence from acquisition and processing experiments. In the first language acquisition of at least the richly inflecting languages Polish and Lithuanian, morphonotactic clusters are acquired earlier than phonotactic clusters (Zydorowicz 2010, Kamandulytė-Merfeldienė 2015). And at least in certain psycholinguistic experiments (cf. the other contributions to this volume), morphonotactic clusters are processed more quickly than phonotactic ones. Therefore, the claim that morphonotactic clusters are more dispreferred than phonotactic clusters should be dropped.

This conclusion is also supported by the ease of diachronic introduction of new, i.e. morphonotactic clusters into languages that lacked them.

A further finding on diachrony is that we have found in German, in analogy to what has been found in other languages, examples of the lexical development of morphonotactic clusters into phonotactic ones because of morphosemantic opacity leading to the loss of morpheme boundaries, as in *Brunst* 'ardour, lust' no longer being related to its former verb base *brenn-en* 'burn', except metalinguistically (cf. Dressler et al. 2019)

Similarly to many other languages, quadruple clusters can be reduced in casual speech. Thus, the normal pronunciation of 2nd SG. *wäsch-st* '(you) wash' is [vɛʃt]. These instances are fairly regular if the NAD distance is minimal, as in this case.

Probably, segmentally identical phonotactic and morphonotactic clusters have different vowel durations (cf. Plag 2014; Zimmerer, Scharinger & Reetz 2014), but it is, as yet, unclear whether these differences lie above the threshold of perceptibility. Moreover, other studies contradict these findings (see the discussion in Leykum & Moosmüller, this volume). In any event, Plag is right in objecting to linguistic models which crucially contain a flow-chart from one submodule to another in a way which presupposes bracket erasure (also criticized in Brown & Hippisley 2012: 273). Our model of morphonotactics (Dressler & Dziubalska-Kołaczyk 2006; Dressler et al. 2010; Korecky-Kröll et al. 2014) does not presuppose such bracket erasure. This also fits Slovak word-medial patterns: assuming that in a flow-chart, inflectional morphology follows derivational morphology, the derivational boundary in *potok* 'stream' must not be erased in order to prevent vowel deletion in Gen.SG. *po-tok-a/u*, in contrast to the deletion of the second vowel in the oblique cases of *ist-ok* 'source' and *otec* 'father' (Dressler et al. 2015).

For results regarding NAD calculations, see section 2.

4.2. Typological conclusions

Phonotactic asymmetries between word-initial, word-final and word-medial positions are well known. This starts with how the universal preference for CV structures (Dziubalska-Kołaczyk 2002, 2009) is realized in the three positions and depending on whether a word is monosyllabic, disyllabic or polysyllabic.

What is interesting for the typological characterization of German is the much greater variety and complexity of word-final than of word-initial clusters, e.g. in contrast to Slavic languages, Latin, Greek and other Indo-European languages. This asymmetry is also reflected in greater type and token frequencies for word-final than for word-initial obstruent clusters.

Type frequency asymmetries proved to be radicalized in token frequency differences, which means that the dominant patterns are more profitable.

This asymmetry has two sources: on the one hand, we have the diachronic result of prehistoric or early historic major vowel deletions in German word-final positions as opposed to the optimal preservation of vowels in word-initial positions. Those lost vowels of word-final syllables were all unstressed, which was not the case for word-initial syllables. On the other hand, we have the more important consequence of German having many short derivational and inflectional suffixes which are monoconsonantal or biconsonantal. But due to the restriction of morphological consonantism to very few consonants, already identified by Jakobson (1962: 108) for Indo-European languages, in German we find only final morphonotactic clusters ending in *-t, -s, -st, -t͡s*. Therefore, it seems a paradox that we find a still more radical restriction for final phonotactic clusters, namely to *-t, -st* and to nouns. The reason is again diachronic: all the final phonotactic nominal triple clusters go back or seem to go back to morphonotactic clusters with a final suffix now ending in *-t* due to the loss of unstressed vowels that followed them or a *-t* added secondarily in early New High German as a phonological addition, as in *Werft* 'shipyard', *Axt* 'axe', *Obst* 'fruit', *sonst* 'otherwise', dialectal *Senft* 'mustard' (Kluge & Götze 1957 sub vocibus).

Word-internally, the contrast between exclusively morphonotactic and exclusively phonotactic triconsonantal clusters seems to be even bigger. Also, here most triconsonantal clusters with two obstruents are only morphonotactic. An among ambiguous consonant clusters, the frequencies of morphonotactic clusters seem to be higher than those of phonotactic clusters. For efficient calculation of these frequency relations, new text-technological tools must be developed.

The fact that in German peripheral positions the NAD preferences for consonant clusters are identical irrespective of whether the most peripheral consonant is included or excluded in the NAD calculations, seems to be specific for Germanic languages. When we checked peripheral consonant clusters in Polish and English according to the list of clusters in Zydorowicz et al. (2016), we found that the (dis)preferredness of consonant clusters is different in Polish depending on whether the most peripheral consonants are included or excluded, but not in English.

Polish and at least Slovak among other Slavic languages (Dressler et al. 2015) differ from German and English with regard to peripheral triple consonant clusters in the following features, which appear to be relevant for the impact of the most peripheral consonant on cluster preferences when they are added to the more interior double consonant clusters:

First of all, the two Slavic languages are consonantal languages to a higher extent than the two Germanic languages. They have a much higher number of different triple consonant clusters than the two Germanic languages. For example, Polish has more than a hundred word-initial triple clusters, German only eight.

Second, Polish has many more word-initial triple morphonotactic clusters in tokens than phonotactic clusters; the two Germanic languages have no word-initial morphonotactic clusters.

Third, for word-final triple consonant clusters, the two Germanic languages have many more morphonotactic than phonotactic clusters, all of them due to the morphological operation of suffixation (i.e. addition). Polish and Slovak have only word-final morphonotactic clusters created through the subtractive morphological operation of deletion of the word-final stem vowel in the genitive plural, e.g. in Pol. *zemst* vs. Nom.SG. *zemsta* 'revenge', Slov. *pomst* vs. Nom.SG. *pomsta* 'revenge'. In addition, Polish and other Slavic languages also create word-initial and word-medial consonant clusters due to vowel deletion in inflection and derivation, as in Pol. Gen.SG. *ps-a* from *pies* 'dog'. German has only rare word-medial cases (see section 3).

Fourth, the most peripheral German consonants in triple consonant clusters in a word-initial position are only /s/ and /ʃ/ (in English only /s/), whereas Polish and Slovak also have many other consonants in this position. In word-final position the most peripheral consonants in German are only /t, s, t͡s/, in English /t, d, s, z/. These consonants are also the preferred final consonants in double clusters. By contrast, many different final consonants occur in Polish and Slovak word-final clusters. Thus, it seems that in the case of strong restrictions on the selection of the most peripheral consonants, the selection is natural, in the sense of not changing the (dis)preferredness of the interior consonant clusters to which they are added. This is reminiscent of those phonotactic analyses which assume for German, as for many other languages, that any third consonant in a tautosyllabic consonant cluster is extrasyllabic or extrametrical (see Wiese 1988, 2000).

This may also explain why, in the diachronic development of German, /t/ was sometimes added to a word-final consonant, as in *Axt* 'axe', *Palast* 'palace', *Obst* 'fruit' from MHG *obes*, *Sekt* 'sparkling wine' from Fr. *vin sec*, dialectal Austrian German *Senft* ← *Senf* 'mustard'.

4.3. Considerations on working with large electronic corpora

Working with large electronic corpora allows us to arrive at more reliable quantitative results. Here, the type-token ratio is very low for all triple clusters. For quadruple clusters we found (see section 2.1) distinct groupings within the whole range from 0.01% to 100%. Thus, the numerically most complex clusters behave differently than the less complex and more numerous triple clusters. The largest subgroup of quadruple clusters has a similar TTR distribution to the triple ones and contains the only four clusters which also include a small phonotactic minority. The more numerous groups of quadruple clusters are only morphonotactic: this again indicates the marked character of complex consonant clusters.

Our corpus-based study relied on the huge electronic corpus AMC, which may be the most complete print media corpus for any nation. This enhanced reliability for quantitative generalizations about the distribution of morphological and lexical patterns of consonant clusters. The disadvantage that such big corpora include many erroneous types of words was at least partially corrected for by manual exclusion of errors and by the restriction to types which have at least 5 tokens in the corpus. We included clusters with fewer than 5 tokens only if the cluster would otherwise not have been represented in our description. In discussions with other native speakers of German we could not think of any potential morphonotactic cluster which does not occur in the AMC.

Clearly new automatic tools should be developed for reducing the error-prone nature of large electronic corpora. More efficient tools are also needed for pattern searches, as we ascertained when studying word-internal clusters.

Even with better tools, the evidence from such an electronic corpus of written adult and adult-directed speech must be considered with caution. The AMC represents just one genre, and it has been found, at least for Modern Greek and Balto-Slavic languages (Dressler et al. 2017) that the distribution of lexical and morphological patterns may differ significantly for different genres.

REFERENCES

Aitchison, Jean (2003) *Words in the Mind: An Introduction to the Mental Lexicon.* 3rd ed. Oxford: Wiley-Blackwell.

Bauer, Laurie (2001) *Morphological Productivity.* Cambridge: Cambridge University Press.

Berg, Thomas (2014) On the relationship between type and token frequency, *Journal of Quantitative Linguistics* 21(3), 199–222.

Bertinetto, Pier Marco; Scheuer, Sylwia; Dziubalska-Kołaczyk, Katarzyna & Agonigi, Maddalena (2006) Intersegmental cohesion and syllable division in Polish. Extended version. *Quaderni del Laboratorio di Linguistica* 6. Pisa: Scuola Normale Superiore.

Blevins, Juliette (2007) The importance of typology in explaining recurrent sound patterns, *Linguistic Typology* 11, 107–113.

Brown, Dunstan & Hippisley, Andrew (2012) Network Morphology. Cambridge: Cambridge University Press.

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

Calderone, Basilio; Celata, Chiara & Laks, Bernard (2014) Theoretical and empirical approaches to phonotactics and morphonotactics, Introduction to special volume on phonotactics and morphonotactics, *Language Sciences* 46, 1–5.

Donohue, Mark; Hetherington, Rebecca; McElvenny, James & Dawson, Virginia (2013) *World Phonotactics Database*. Canberra: Department of Linguistics, Australian National University.

Dressler, Wolfgang U. (1984) Explaining Natural Phonology. *Phonology Yearbook* 1, 29–51.

Dressler, Wolfgang U. (1985) *Morphonology: The Dynamics of Derivation*. Ann Arbor, Mich.: Karoma Publishers.

Dressler, Wolfgang U. (1996a) A functionalist semiotic model of morphonology. In: Singh, Rajendra (ed.) *Trubetzkoy's Orphan*. Amsterdam: Benjamins, 67–83.

Dressler, Wolfgang U. (1996b) Reply to Janda and Walker. In: Singh, Rajendra (ed.) *Trubetzkoy's Orphan*. Amsterdam: Benjamins, 102–105.

Dressler, Wolfgang U. (1999) On a semiotic theory of preferences in language. *Peirce Seminar Papers* 4, 389–415.

Dressler, Wolfgang U. & Dziubalska-Kołaczyk, Katarzyna (2006) Proposing morphonotactics, *Italian Journal of Linguistics* 18, 249–266.

Dressler, Wolfgang U.; Dziubalska-Kołaczyk, Katarzyna & Pestal, Lina (2010) Change and Variation in Morphonotactics, *Folia Linguistica Historica* 31, 51–67.

Dressler, Wolfgang U.; Hliničanová, Miroslava; Ďurčo, Matej; Mörth, Karlheinz & Korecky-Kröll, Katharina (2015) Phonotactic vs. morphonotactic obstruent clusters in Slovak and German, *Italian Journal of Linguistics* 27, 45–60.

Dressler, Wolfgang U.; Ketrez, F. Nihan & Kilani-Schoch, Marianne (eds) (2017) *Nominal Compound Acquisition*. Amsterdam: Benjamins.

Dressler, Wolfgang U. & Kilani-Schoch, Marianne (2017) Natural Morphology. In: Hippisley, Andrew & Stump, Gregory (eds) *The Cambridge Handbook of Morphology*. Cambridge: Cambridge University Press, 356–389.

Dressler, Wolfgang U.; Kononenko, Alona; Sommer-Lolei, Sabine; Korecky-Kröll, Katharina; Zydorowicz, Paulina & Kamandulytė-Merfeldienė, Laura (2019) Morphological richness and transparency and the genesis and evolution of morphonotactic patterns, *Folia Linguistica Historica* 40(1), 85–106. doi:10.1515/flih-2019-0005

Dressler, Wolfgang U.; Libben, Gary & Korecky-Kröll, Katharina (2014) Conflicting vs. convergent vs. interdependent motivations in morphology. In: MacWhinney, Brian; Malchukov, Andrej & Moravcsik, Edith (eds) *Competing Motivations in Grammar and Usage*. Oxford: Oxford University Press, 181–196.

Du, Lifang & Zhang, Xuezhong (2010) A survey of the measurements of morphological productivity, *English Language Teaching* 3, 60–63.

Dziubalska-Kołaczyk, Katarzyna (2002) *Beats-and-Binding Phonology*. Frankfurt: Lang.

Dziubalska-Kołaczyk, Katarzyna (2009) NP extension: B&B phonotactics, *PSiCL* 45, 55–71.

Dziubalska-Kołaczyk, Katarzyna (2014) Explaining phonotactics using NAD, *Language Sciences* 46, 6–17.

Dziubalska-Kołaczyk, Katarzyna (2019) On the structure, survival and change of consonant clusters, *Folia Linguistica Historica* 53(s40-1), 107–127.

Dziubalska-Kołaczyk, Katarzyna; Pietrala, Dawid & Aperliński, Grzegorz (2014) *The NAD Phonotactic Calculator – an online tool to calculate cluster preference in English, Polish and other languages*. <http://wa.amu.edu.pl/nadcalc/>.

Dziubalska-Kołaczyk, Katarzyna & Weckwerth, Jarosław (2002) *Future Challenges for Natural Linguistics*. Munich: Lincom.

Fehringer, Carol (2011) Allomorphy in the German genitive. A paradigmatic account. *Zeitschrift für germanistische Linguistik* 39, 90–112.

Gagné, Christina (2009) Psycholinguistic Perspectives. In: Lieber, Rochelle & Štekauer, Pavol (eds) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 255–271.

Heidolph, Karl Erich; Flämig, Walter & Motsch, Wolfgang (eds) (1981) *Grundzüge einer deutschen Grammatik* (Main features of German grammar). Berlin: Akademie-Verlag.

Hirsch-Wierzbicka, Ludomira (1971) *Funktionelle Belastung und Phonemkombination am Beispiel einsilbiger Wörter der deutschen Gegenwartssprache* (Functional stress and phoneme combination by the example of monosyllabic words of contemporary German). Hamburg: Buske.

Hongbo, Ji; Gagné, Christina L. & Spalding, Thomas L. (2011) Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds, *Journal of Memory and Language* 65, 406–430.

Hyman, Larry (2007) Where's Phonology in Typology? *Linguistic Typology* 11, 265–271.

Hyman, Larry & Plank, Frans (eds) (2018) *Phonological Typology*. Berlin: De Gruyter.

Jakobson, Roman (1962) *Selected Writings. Vol. I. Phonological Studies.* The Hague: Mouton.

Jespersen, Otto (1904) *Lehrbuch der Phonetik* (Textbook of Phonetics). Leipzig: B.G. Teubner.

Kamandulytė-Merfeldienė, Laura (2015) Morphonotactics in L1 acquisition of Lithuanian: TD vs SLI. *Eesti Rakenduslingvistika Ühingu aastaraamat* (Estonian papers in applied linguistics) 11, 95–109.

Kilani-Schoch, Marianne & Dressler, Wolfgang U. (2005) Morphologie naturelle et flexion du verbe francais (Natural morphology and French verb inflection). Tübingen: Narr.

Kluge, Friedrich & Götze, Alfred (1957) *Etymologisches Wörterbuch der Deutschen Sprache* (Etymological dictionary of the German language). Berlin: de Gruyter.

Korecky-Kröll Katharina; Dressler, Wolfgang U.; Freiberger, Eva M.; Reinisch, Eva; Mörth, Karlheinz & Libben, Gary (2014) Morphonotactic and phonotactic processing in German-speaking adults, *Language Sciences* 46, 48–58.

Lang, Ewald & Zifonun, Gisela (ed.) (1996) *Deutsch – typologisch* (German – typological). Berlin: de Gruyter.

Libben, Gary (1998) Semantic transparency in the processing of compounds: consequences for representation, processing, and impairment, *Brain and Language* 61, 30–44.

Luschützky, Hans Christian (1992) *Zur Phonologie der Affrikaten* (The phonology of affricates). Frankfurt: Hector.

Maddieson, Ian (2006) Correlating phonological complexity: data and validation, *Linguistic Typology* 10, 106–123.

Maddieson, Ian (2013) Consonant Inventories. In: Dryer, Matthew S. & Haspelmath, Martin (eds) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/1> (23 April 2018).

Mathesius, Vilém (1928) On linguistic characterology with illustration from modern English, *Actes du Premier congres international de linguistes à la Haye du 10–15 avril* 1928: 56–63.

McEnery, Tony & Hardie, Andrew (2012) *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.

Meinhold, Gottfried & Stock, Eberhard (1980) *Phonologie der deutschen Gegenwartssprache* (Phonology of contemporary German). Leipzig: Bibliographisches Institut.

Ohala, John J. (1990) The phonetics and phonology of aspects of assimilation. In: Kingston, John & Beckman Mary E. (eds) *Papers in Laboratory Phonology I*. Cambridge: Cambridge University Press, 258–275.

Oppermann, Johannes (2004) Zurück zu den Wurzeln (Back to the roots). MA thesis, University of Vienna.

Orzechowska, Paula & Wiese, Richard (2011) Reconstructing the Sonority Hierarchy, *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, China, 1542–1545.

Orzechowska, Paula & Wiese, Richard (2015) Preferences and variation in word-initial phonotactics: A multi-dimensional evaluation of German and Polish, *Folia Linguistica* 49(2): 439–486.

Plag, Ingo (2014) Phonological and phonetic variability in complex words: An uncharted territory, *Rivista di Linguistica* 26, 209–228.

Ransmayr, Jutta; Mörth, Karlheinz & Matej, Ďurčo (2017) AMC (Austrian Media Corpus) – Korpusbasierte Forschungen zum österreichischen Deutsch (Corpus-based research on Austrian German). In: Resch, Claudia & Dressler, Wolfgang U. (eds) *Digitale Methoden der Korpusforschung in Österreich* (Digital methods of corpus research in Austria). Wien: Verlag der Österreichischen Akademie der Wissenschaften, 27–38.

Shosted, Ryan (2006) Correlating complexity: a typological approach, *Linguistic Typology* 10, 1–40.

Sievers, Eduard (1876) *Grundzüge der Lautphysiologie zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen* (Fundamentals of sound physiology as an introduction to the study of phonology of Indo-European languages). Leipzig: Breitkopf und Hartel.

Szczepaniak, Renata (2010) Während des Flug(e)s/des Ausflug(e)s (During the flight/the excursion): German Short and Long Genitive Endings between Norm and Variation. In: Lenz, Alexandra N. & Plewnia, Albrecht (eds) *Grammar between Norm and Variation*. Frankfurt: Lang, 103–126.

Tomasello Michael (2003) *Constructing a Language: A Usage-based Theory of Language Acquisition*. Cambridge: Harvard University Press.

Trubetzkoy, Nikolaj S. (1939) *Grundzüge der Phonologie* (Main features of phonology). Prag: Travaux du Cercle Linguistique de Prague, 7.

Whitney, William D. (1865) The relation of vowel and consonant, *Journal of the American Oriental Society* 8, 277–300.

Wiese, Richard (1988) *Silbische und Lexikalische Phonologie: Studien zum Chinesischen und Deutschen* (Syllabic and lexical phonology: Studies on Chinese and German). Tübingen: Niemeyer.

Wiese, Richard (1991) Was ist extrasilbisch im Deutschen und warum? (What is extrasyllabic in German and why?), *Zeitschrift für Sprachwissenschaft* 10, 112–133.

Wiese, Richard (2000) *The Phonology of German*. Oxford: Oxford University Press.

Zimmerer, Frank; Scharinger, Mathias & Reetz, Henning (2014) Phonological and morphological constraints on German /t/-deletions, *Journal of Phonetics* 45, 64–75.

Zydorowicz, Paulina (2010) Consonant clusters across morpheme boundaries: Polish morphonotactic inventory and its acquisition, *PSiCL* 46, 565–588.

Zydorowicz Paulina; Orzechowska, Paula; Jankowski, Michał; Dziubalska-Kołaczyk, Katarzyna; Wierzchoń, Piotr & Pietrala, Dawid (2016) *Phonotactics and morphonotactics of Polish and English. Theory, description, tools and applications*. Poznań: Adam Mickiewicz University Press.

Appendix 10. Research article 4: Kononenko-Szoszkiewicz, 2023)

“(Mor)phonotactics of Ukrainian. The study of word-initial consonant clusters”, 2023, [accepted for publication in Italian Journal of Linguistics].

(Mor)phonotactics of Ukrainian. The study of word-initial consonant clusters

Alona Kononenko-Szoszkiewicz,
Faculty of English, Adam Mickiewicz University, Poznan, Poland, <alokon1@amu.edu.pl>

The present paper aims to provide the first analysis of Ukrainian phonotactics and morphonotactics, compare them qualitatively and quantitatively, and explain the difference between these two perspectives. Further, the paper explores the morphological complexity of consonant clusters in the Ukrainian language. The research is limited to consonant clusters in word-initial position compared to earlier studies in other Slavic languages, namely Russian and Polish. With respect to markedness, two hypotheses were tested, suggesting that morphonotactic clusters are expected to be less preferred than phonotactic, and that cluster preferability is directly proportional to frequency. Additionally, there have been discussed predictions of clusters' preferability derived from the Net Auditory Distance principle.

KEYWORDS: UKRAINIAN, PHONOTACTICS, MORPHONOTACTICS, CONSONANT CLUSTERS, SLAVIC LANGUAGES

1. *Introduction*

Ukrainian is spoken by more than 35 million people around the globe (Lewis *et al*. 2016). This number is likely to grow as language learning applications observe the unprecedented interest in acquiring Ukrainian (von Ahn, 2022). Nevertheless, it remains one of the least investigated languages of the Slavic language family in terms of phonetics and phonology. Most publications on Ukrainian phonetics date back to the 1970s or earlier and do not represent the state of the modern language and present-day investigations. As Vakulenko (2018) highlighted, the central issue of Ukrainian phonetics is that contemporary judgments about the language are based on outdated phonetic material obtained from just one speaker and processed with old-fashioned phonetic methods.

However, there are a few recent descriptions of the Ukrainian phonetic system, e.g., by Buk *et al*. (2008) and Pompino-Marschall *et al.* (2016). Yet, they have been heavily criticized by Vakulenko (2019) due to the lack of relevant experimental material and coherent explanations of the assumptions. Thus, the question of phonetic realizations of variations within the modern Ukrainian language remains open and heavily depends on various regional dialects. According to the "Atlas of the Ukrainian language," there are three major dialects that are characterized by phonetic, lexical and grammatical distinctions (Matvijas *et al.* 2001). Most publications dedicated to Ukrainian phonetics present or only briefly mention a selected group of phonemes in their syntagmatic organization, but phonotactics (not to speak of morphonotactics) has never been a subject of a study.

1.1. *Ukrainian phonotactics*

The monograph "Contemporary standard Ukrainian. Phonetics" (Bilodid, 1969) remains one of the most significant works in Ukrainian phonetics, presenting experimental data on consonants. Although there is no separate chapter dedicated to Ukrainian phonotactics, the author analyzes some frequent combinatory possibilities of Ukrainian phonemes based on the texts of various literary genres. This study's methodology relied on counting frequency of occurrence of phonemes with a view to differences in voicing, manner of articulation, place of articulation, and soft vs. hard consonant opposition. As concluded by the author, the Ukrainian language prefers the following combinations of consonants: plosive + sonorant, fricative + plosive, fricative + affricate, and fricative + sonorant, rather than combinations in which these groups of phonemes occur in the reverse order. Nevertheless, neither examples nor quantitative information regarding the inventory of consonant clusters (CC) were provided.

A monograph, “The History of Ukrainian language. Phonetics” by Zhovtobriuch (1979) outlines combinatorial possibilities of consonants clustered together. The author described only the possible combinations of plosive + sonorant, fricative + sonorant, voiced fricative + voiced plosive, voiceless fricative + voiceless plosive, affricate + fricative, bilabials + lateral, affricate + fricative, two sonorants. Among sequences of three consonants, the author mentioned just combinations of /z/ and /s/ followed by plosives /d/, /t/, /k/. Quadruple Ukrainian consonant clusters were not mentioned at all. In another publication on the phonetic description of Ukrainian by Zilynski (1979), the author mentioned possible combinations of two stops, stops + fricatives, and sequences of homorganic consonants. Thus, there are a few descriptions of Ukrainian phonotactics, but the information remains scattered and incomplete. However, there is no publication which would present a comprehensive picture of the phonotactic and morphonotactic inventory of the Ukrainian language.

As a rule, the division into vocalic vs. consonantal languages could be distinguished according to the number of vocalic and consonantal elements in the phonemic inventories or by syllable structure and the number of consonant clusters. According to Isachenko (1963), a phonemic opposition between plain and palatalized consonants across different articulation classes implies the consonantal character of the Ukrainian language. The inventory of consonants compared to the number of vowels in the Ukrainian phonemic system constitutes 72%, while Polish has 87.5%, which is the highest ratio among all Slavic languages (Majewicz, 1989). Such classification is connected with the syllabic patterns occurring in particular languages: open syllables are characteristic of the vocalic type, where the CV and V syllables predominate, the V syllables being relatively frequent. In the languages of the intermediate type, syllables closed by a single consonant additionally occur, the CV syllables being the most frequent. Closed syllables and rich consonant clusters are characteristic of the consonantal type (Majewicz, 1989). According to these criteria, all Slavic languages could be characterized as consonantal. Yet the degree of consonantism and the number of consonant clusters present in a language signify gradual typological differences.

According to Zilynskyj (1979), Ukrainian generally does not tolerate long clusters of consonants, and secondary syllables are formed with sonorant consonants. It either completely eliminates them by dropping the sonorant or turns them into syllables with full voice by inserting a vowel. For instance, the Polish language accepts all kinds of combinations of sonorant and obstruent: SO, OS, and OSO in initial, final, and medial positions (e.g., *wiatr* ‘wind’, *rwać* ‘to tear apart’, *brda* ‘beard’, etc. pronounced with non-syllabic [r]). The same situation is found in the Sorbian languages, but also in Russian and Ukrainian. Still, in these languages, the frequency of the initial SO- and final -OS clusters containing non-syllabic sonants is lower than in Polish (Sawicka, 2001).

A syllable structure of Ukrainian has been analyzed by Czaplicki (2007) from the Optimality Theory perspective (Prince and Smolensky, 1993). The author described selected consonant clusters in word-initial, medial and final positions according to the Sonority Sequencing Principle. Another way to analyze consonant clusters could be from the perspective of markedness (Eckmann, 1977). In the markedness approach, when applied to onsets and codas, it is considered that the longer the onsets and codas are, the more marked they are. With regards to morphonotactics, it has been generally hypothesized that morphonotactic sequences are more likely to be marked, therefore, dispreferred (Dressler & Dziubalska-Kołaczyk, 2006).

1.2. Ukrainian morphonotactics

The distinction between morphonotactics and phonotactics has been introduced by Dressler & Dziubalska-Kołaczyk (2006). While phonotactics studies permissible combinations of consonants clustered together, morphonotactics refers to the combinations of consonants that appear only at morpheme boundaries. Thus, the consonant cluster /dv-/ as in *dva* ‘two’ is considered phonotactic or lexical, but the consonant cluster /z+ts/ as in *z+cilyty* ‘to heal’ comes into being through adding a prefix to the following consonant, therefore it is morphonotactic. However, some consonant clusters can occur both in phonotactic and morphonotactic combinations. For instance, /vl-/ in *vlada* ‘power’ is phonotactic since the initial phoneme /v/ is part of a word root, but in *v+lazyty* ‘to get in’ it is morphonotactic because v- is a prefix.

Over the previous ten years, an array of scholarly investigations has emerged, addressing various facets of morphonotactics within different domains of linguistics, such as language acquisition, psycholinguistics, corpus linguistics, and typological studies. Predominant languages of research on morphonotactics represent different language families, such as:

- Slavic, e.g., Slovak (Dressler & Hliničanová, 2015); Polish (Zydorowicz et al. 2016), Russian (Dressler & Kononenko-Szoszkiewicz, 2019), Croatian (Kelić & Dressler, 2019),
- Baltic, e.g., Lithuanian (Kamandulytė-Merfeldienė, 2015);
- Romance, e.g., Italian (Dressler & Dziubalska-Kołaczyk, 2006), French (Köpke et al. 2021);
- Germanic e.g., German (Korecky-Kröll et al. 2014), English (Zydorowicz et al. 2016).

Typological differences here are of prior interest because the languages with a richer morphology, predominantly Slavic languages, are supposed to have more morphonotactic consonant clusters. For instance, Polish can tolerate up to four-segment initial cluster as in /v+z+gl-/ *wzgledny* 'relative' and maximum of five consonants in word-final position as in /-mpstf/ *przestępstw* 'crimes' (only in Gen. case). Thus, this pioneering work on Ukrainian phonotactics could be a starting point for future comparative typological studies.

For the purpose of the present research, an alternative approach for cluster evaluation was applied based on the universal model of phonotactics constructed within the Beats-&-Binding phonology model (Dziubalska-Kołaczyk, 2002, 2009). Such a choice is motivated by the fact that this model goes beyond purely sonority-based models and is not attached to any of the traditional syllabification models. The model presents syllabic nuclei as beats and consonants bound to them but does not assume syllabic boundaries. By taking into account the perceptual contrast between beats and non-beats it allows to evaluate cluster preferability and to establish a hierarchy of the preferences of clusters from the most preferred (unmarked) to the least preferred (marked). Perceptual contrast of the consonants is measured employing the Net Auditory Distance principle (NAD) (Dziubalska-Kołaczyk 2009, 2014). A new model of NAD is not only based on the sonority balance between the phonemes but also includes manner of articulation, place of articulation as well as sonorant-obstruent distinction. By means of an online tool - the NAD phonotactic calculator (Dziubalska-Kołaczyk et al. 2007, 2014), there has been established a hierarchy of preferences for Ukrainian word-initial consonant clusters including the division of phonotactic and morphonotactic CC.

A major source of morphonotactic clusters in Ukrainian is derivation. According to the "Dictionary of affixal morphs of Ukrainian", there are 145 prefixal morphs. Moreover, 43 prefixes were borrowed into Ukrainian from other languages (*a-, ad-, ab-, ana-, anti-, apo-, archi-, hyper-, hypo-, de-, dis-, dia-, e-, ek-, eks-, extra-, en-, epi-, in-, inter-, intro-, infra-, ipo-, kata-, kon-, ko-, kontr-, meta-, par-, para-, per-, peri-, post-, pre-, pro-, re-, sin-, sub-, super-, sur-, trans-, ultra-*). Ten units belong to the complex, secondary prefix combinations: *za+v-, z+ne-, na+v-, ne+do-, o+bez-, po+za-, po+nad-, po+pid-, s+piv-, s+pid-* (Klimenko et.al, 1998). There are two productive prefixes z- (also assimilated as s-) and v-, which give rise to the establishment of morphonotactic consonant clusters. The Old-Russian prefixes *sъ-* 'off; with' and *jьz-* 'out of' have merged into a single prefix - modern Ukr. z- (Andersen, 1969). The prefix z- also occurs as preposition, but before voiceless consonants (/k/, /p/, /t/, /h/) due to voice assimilation, it is pronounced as /s/. Such pronunciation has also been reflected in Ukrainian orthography, e.g. *s+pytaty* 'to ask', *s+xodyty* 'to go', *s+kazaty* 'to say'. When it appears in nouns, it has two semantic sources, one meaning 'together', 'with' and the other 'from', 'out of', and in verbs it occurs as a marker of perfective aspect. Yet formations of verbs in which z- serves as perfectivizing element may have the original sense of the prefix obscured (Press & Pugh, 2015).

The non-syllabic consonantal prefix v- is the most productive in verb formation. The meaning of the verb prefixes v- (also vi-, u-, u+vi-) is ambiguous and can convey various meanings. For instance, it is a special-objective as in *v+bigaty* 'to run in', time-objective as in *v+topyty* 'to drown', it can also signify an effect as in *v+movyty* 'to persuade, lit. to say to'. Thus, all Ukrainian morphonotactic consonant clusters are derived due to the prefixation of s-, z-, v- attached word-initially.

1.3. *Data and methodology*

The corpus linguistic research is based on the data extracted from The General Regionally Annotated Corpus of Ukrainian (UA: Генеральний регіонально анотований корпус української мови by Shvedova et.al 2017-2022). The corpus design has been inspired by the model of existing reference corpora such as Czech, Russian, or Polish national corpora, and the British National Corpus. This is the first and so far, the only corpus of the Ukrainian language which contains texts annotated by regional markup. The corpus encompasses the timespan between 1816 and 2022 and includes over 90 thousand texts of different genres by about twenty-six thousand authors. For present research analysis there has been used the GRAC-14 version of the corpus which encompasses about 860 million tokens. Running the corpus query language (CQL) operations allowed to automatically generate a list of word types containing a specific consonant cluster along with its frequency in the corpus. During the data selection process, different lemmas of the same word have been counted as one-word type. The word type count has been limited to words with at least five tokens.

2. *Results*

2.1. *Word-initial double consonant clusters*

Double consonant clusters constitute the largest group of word-initial consonant clusters in Ukrainian. There are 112 word-initial consonant clusters (Table 1). The table below represents the combinatorial inventory of word-initial double consonant clusters. Based on previous assumptions of Bilodid (1969), the data from the corpus confirmed that the most frequent combination according to lemma type is a stop followed by a sonorant. There are overall 23 consonant clusters of that type. The three lexical clusters /pr-/, /kr-/, /tr-/ represent the most frequent consonant combinations in the corpus. Also, the three most frequent triple consonant clusters begin with the voiced velar stop /g/ and four consonant clusters begin with the voiced glottal fricative /ɦ/.

		b	ʋ	ɦ	g	d	ʒ	d͡z	d͡ʒ	z	k	l	m	n	p	r	s	t	f	x	ts	t͡ʃ	ʃ
IPA	Ukr	б	в	г	ґ	д	ж	дз	дж	з	к	л	м	н	п	р	с	т	ф	х	ц	ч	ш
b	б	-	-	-	-	-	-	-	+	-	-	+	-	-	-	+	-	-	-	-	-	-	-
ʋ	в	+	-	+	-	+	+	-	-	+	+	+	+	+	+	+	+	+	-	+	+	+	+
ɦ	г	-	+	-	-	-	-	-	-	-	-	+	-	+	-	+	-	-	-	-	-	-	-
g	ґ	-	+	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-
d	д	-	+	-	-	-	-	-	-	-	-	-	+	+	-	+	-	-	-	-	-	-	-
ʒ	ж	-	+	-	-	-	-	-	-	-	-	+	+	+	-	+	-	-	-	-	-	-	-
d͡z	дз	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
d͡ʒ	дж	-	-	+	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
z	з	+	+	+	-	+	+	-	-	-	-	+	+	+	-	+	-	-	-	-	+	+	+
k	к	-	+	-	-	-	-	-	-	-	-	+	-	+	-	+	+	-	-	-	-	-	-
l	л	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-
m	м	-	-	-	-	-	-	-	-	-	-	+	-	+	-	+	-	-	-	-	-	+	-
n	н	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
p	п	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	+	-	+	-	-	+
r	р	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
s	с	-	+	-	-	-	-	-	-	-	+	+	+	+	+	+	-	+	+	+	+	-	-
t	т	-	+	-	-	-	-	-	-	-	+	+	+	-	-	+	-	-	-	+	-	-	-
f	ф	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-
x	х	-	+	-	-	-	-	-	-	-	-	+	+	+	-	+	-	+	-	-	-	-	-
t͡s	ц	-	+	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-
t͡ʃ	ч	-	+	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	+	-	-	-

ʃ	ш	-	+	-	-	-	-	-	-	-	+	+	+	+	+	+	-	+	-	+	-	-	-

Table 1. *Combinatory possibilities of Ukrainian phonemes*

The list of word-initial double consonant clusters is provided in the Appendix (Table 5). All clusters are exemplified by the most frequent lemma type in the corpus, transliterated, translated into English, specified by the type of clusters, i.e. phonotactic, morphonotactic or both. Among 112 word-initial clusters, the majority of clusters are phonotactic – 81, six consonant clusters are exclusively morphonotactic with no lexical counterparts: /vt-/, /vʒ-/, /vx-/, /zʒ-/, /zʃ-/, /vts-/. Eighteen consonant clusters occur both as morphonotactic and phonotactic, namely /sp-/, /st-/, /sk-/, /zn-/, /zv-/, /zm-/, /zd-/, /zb-/, /vs-/, /zr-/, /vn-/, /vr-/, /vl-/, /vp-/, /sx-/, /vz-/, /vd-/, /vt͡ʃ-/. For instance, *sp-* as in sp+osib 'a way' (phonotactic) but s+pytaty 'to ask' (morphonotactic).

2.2. Word-initial triple consonant clusters

There are less than half of word-initial triple consonant clusters as doubles. The overall number of triple clusters is 69, out of which 52 clusters are morphonotactic (Table 6). Only three consonant clusters, namely /spr-/, /zbr-/, /zɦr-/ occur both as morphonotactic and phonotactic: /s+pr/ in *s+prava* 'business', /spr/ in *sprytny* 'agile' /z+br/ in *z+brehaty* 'to lie', /zbr/ in *zbroya* 'weapon', /z+ɦr/ in z+gribaty 'to shovel, /zɦr/ *zgraya* 'flock'.

2.3. Word-initial quadruple consonant clusters

Ukrainian allows strings of four phonemes in an initial position. Thus there are some word-initial quadruple clusters in Ukrainian such as /vzdr/ in the dialectal perfective verb *v+z+driv* 's(h)e has seen', /vpxn/ in the vocative case *v+pxny* 'shove sth in', in the dialectal perfective verb /vstr/ in *v+striv* 's(h)e has met' similarly to the Standard Ukrainian *zu+strity* 'to meet', /vʃkv/ in *v+škvaryty* 'to strike', /vʃtr/ in *v+štryknuty* 'to prick. All quadruple consonant clusters are morphonotactic due to the morphological concatenation of the prefix /v/ with the following consonants. There is only one quadruple cluster beginning with /s/ as in /sʃkr/ in *s+škrebty* 'to scrape off'.

2.4. The NAD preference

The phonotactic calculator is a software designed by Dziubalska-Kołaczyk et. al. (2007, 2014) for measuring the auditory distances between the neighbouring phonemes as defined by the NAD principle. The calculator allows measuring the preferability of the cluster according to its position in a word (initial, medial or final) as well as to build up the hierarchy of preferability of clusters from the most preferred to the least defined by the NAD product. The NAD product indicates a mean number of all the distances between the neighbouring phonemes in the cluster. It was introduced to the calculator in order to assign a preferability index which is "a number denoting a degree to which a given preference is observed" (Dziubalska-Kołaczyk, 2019).

The settings for English, German and Polish were previously implemented in the calculator, but the parameter values for Ukrainian were not specified. Therefore, the values for Ukrainian were adopted by the author in accordance with the International Phonetic Alphabet. The phonetic description of Ukrainian is based on the illustration of the IPA compiled by Pompino-Marschall *et al*. (2017).

Following the study on Polish and English (Zydorowicz *et al.* 2016) the purpose of the present research is to analyze the phonotactic inventory of Ukrainian regarding the composition of clusters, the degree of preferability and frequency. Hypothesis 1, previously formulated by Dressler & Dziubalska-Kołaczyk (2006), suggests that the degree of phonological preferability is inversely proportional to morphological complexity. Thus, morphonotactic consonant clusters are expected to be less preferred than phonotactic ones. The second hypothesis states that the degree of cluster preferability is directly proportional to frequency. Preferred clusters are expected to be more frequent than dispreferred.

3. *Discussion*

To verify hypotheses, the status of word-initial double clusters has been calculated with the help of the NAD calculator. As demonstrated in the Table2, among 112 word-initial double consonant clusters, 61 clusters are preferred, and 51 are dispreferred.

	IPA	b	ʋ	ɦ	g	d	ʒ	d͡z	d͡ʒ	z	k	l	m	n	p	r	s	t	f	x	t͡s	t͡ʃ	ʃ
IPA	Ukr	б	в	г	ґ	д	ж	дз	дж	з	к	л	м	н	п	р	с	т	ф	х	ц	ч	Ш
b	б	-	-	-	-	-	-	-	NO	-	-	YES	-	-	-	YES	-	-	-	-	-	-	-
ʋ	в	NO	-	YES	-	NO	YES	-	-	NO	YES	NO	NO	NO	NO	NO	NO	NO	-	YES	NO	YES	YES
ɦ	г	-	YES	-	-	-	-	-	-	-	-	YES	-	YES	-	YES	-	-	-	-	-	-	-
g	ґ	-	YES	-	-	-	-	-	-	-	-	YES	-	-	-	YES	-	-	-	-	-	-	-
d	д	-	YES	-	-	-	-	-	-	-	-	-	YES	YES	-	YES	-	-	-	-	-	-	-
ʒ	ж	-	YES	-	-	-	-	-	-	-	-	YES	YES	NO	-	YES	-	-	-	-	-	-	-
d͡z	дз	NO	YES	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
d͡ʒ	дж	-	-	NO	-	-	-	-	-	-	-	-	YES	-	-	-	-	-	-	-	-	-	-
z	з	NO	YES	NO	-	NO	NO	-	-	-	-	YES	YES	NO	-	YES	-	-	-	-	NO	NO	NO
k	к	-	YES	-	-	-	-	-	-	-	-	YES	-	YES	-	YES	NO	-	-	-	-	-	-
l	л	-	YES	-	-	-	-	-	-	-	-	-	-	NO	-	-	-	-	-	-	-	-	-
m	м	-	-	-	-	-	-	-	-	-	-	NO	-	NO	-	YES	-	-	-	-	-	NO	-
n	н	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
p	п	-	-	-	-	-	-	-	-	-	-	YES	-	-	-	YES	-	NO	-	NO	-	-	NO
r	р	-	YES	-	-	-	NO	-	-	-	-	-	-	-	-	-	-	NO	-	-	-	-	-
s	с	-	YES	-	-	-	-	-	-	-	NO	YES	YES	NO	NO	YES	-	NO	NO	NO	NO	-	-
t	т	-	YES	-	-	-	-	-	-	-	NO	YES	YES	-	-	YES	-	-	-	NO	-	-	-
f	ф	-	-	-	-	-	-	-	-	-	-	YES	-	-	-	YES	-	-	-	-	-	-	-
x	х	-	YES	-	-	-	-	-	-	-	-	YES	YES	YES	-	YES	-	NO	-	-	-	-	-
t͡s	ц	-	YES	-	-	-	-	-	-	-	-	-	YES	NO	-	-	-	-	-	-	-	-	-
t͡ʃ	ч	-	YES	-	-	-	-	-	-	-	-	YES	YES	-	-	-	-	-	-	NO	-	-	-
ʃ	ш	-	YES	-	-	-	-	-	-	-	YES	NO	YES	NO	NO	YES	-	NO	-	NO	-	-	-

Table 2. *NAD preferences for word-initial doubles*

Regarding the consonant clusters' inventory, the majority of word-initial double clusters are phonotactic. For word-initial doubles, the data strongly supports Hypothesis 1 since phonotactic consonant clusters are twice as much preferred than dispreferred, also having a high degree of word-type frequency (Table 3).

	Preferred	Dispreferred
Morph	N=4 (ʋk-, ʋʒ-, ʋʃ-, ʋx-)	N=9 (zh-, ʋt-, ʋm-, ʋb-, ztʃ-, zts-, zʒ-, zʃ-, ʋts-)
Phon	N=53 (pr-, kr-, tr-, hr-, br-, dr-, sl-, pl-, sʋ-, bl-, dʋ-, hl-, kl-, xr-, sm-, zl-, fr-, kʋ-, xl-, kn-, tʋ-, xʋ-, hn-, fl-, sr-, gr-, ʃʋ-, ʃl-, tsʋ-, xm-, ʃm-, ʋh-, mr-, dn-, hʋ-, dzʋ-, tʃʋ-, ʒm-, rʋ-, tl-, ʃr-, gl-, gʋ-, tʃm-, ʒr-, tʃl-, ʒʋ-, ʒl-, tsm-, dm-, lʋ-, tm-, xn-)	N=28 (ʃt-, ln-, ʃk-, ʃp-, mn-, sn-, ml-, pt-, sf-, sts-, ks-, bdʒ-, xt-, pʃ-, tk-, ʃn-, rʒ-, px-, ʃx-, ʒn-, tsn-, tx-, dzb-, mtʃ-, dzh-, dʒh-, rt-, tʃx-)
Both	N=4 (zʋ-, zm-, zr-, ʋtʃ-)	N=14 (sp-, st-, sk-, zn-, zd-, zb-, ʋs-, ʋn-, ʋr-, ʋl-, ʋp-, sx-, ʋz-, ʋd-)

Table 3. *NAD preferences of word-initial doubles*

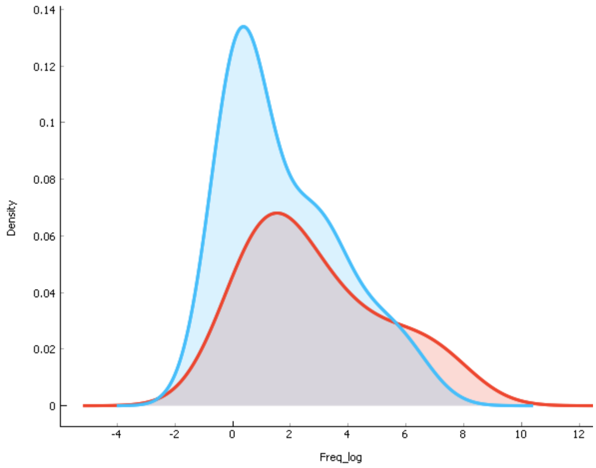
For triple consonant clusters, the majority of clusters are morphonotactic and strongly dispreferred, which again supports Hypothesis 1. The second prediction regarding frequency and cluster preference has been confirmed partially, since generally, there are more dispreferred morphonotactic clusters (N=41) than preferred (N=28). Yet the five most frequent consonant combinations are morphonotactic and preferred. (Table 4).

	Preferred	Dispreferred

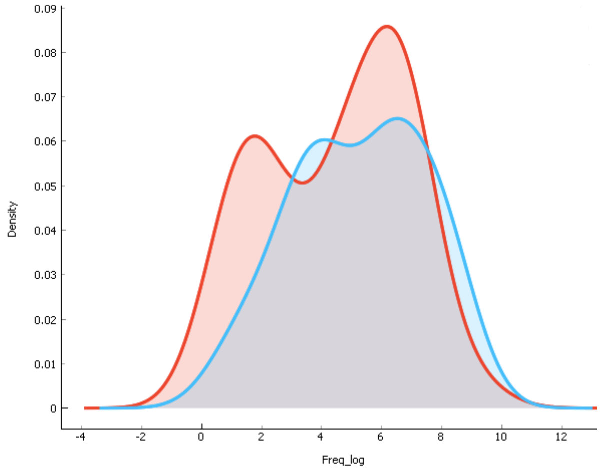
Morph	N=15 (skl-, spl-, zdr-, stʊ-, zbl-, zxl-, sxr-, zdʊ-, zxn-, shl-, stl-, zxʊ-, zdm-, sʧl-, zʤʊ-)	N=37 (ʊst-, ʊpr-, ʊtr-, ʊkr-, ʊkl-, ʊpl-, ʊsp-, ʊxl-, ʊzd-, ʊsl-, ʊsm-, ʊxr-, ʊsk-, ʊzr-, ʊdʊ-, ʊzn-, ʊdr-, ʊxn-, ʊbr-, ʊzʊ-, ʊxʊ-, sʃt-, ʊzl-, ʊzb-, ʊzx-, zmr-, sʃʧ-, ʊbl-, ʊʧp-, ʊdm-, ʊpn-, ʊpx-, ʊtl-, ʊzm-, ʊtn-, ʊxl-, ʊtk-)
Phon	N=10 (str-, skr-, ʃtr-, skʊ-, ʃkr-, skn-, ʃpr-, ʃkl-, sfr-, tkn-)	N=4 (ʊʃʧ-, smr-, pxn-, ʊʧʊ-)
Both	N=3 (spr-, sbr-, zhr-)	

Table 4. *NAD preferences for word-initial triples*

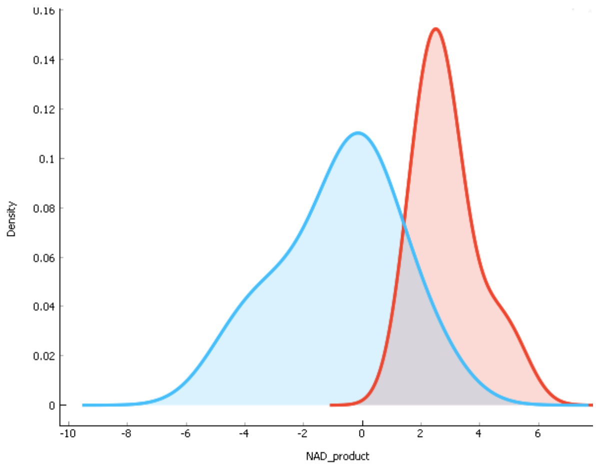
To validate the second hypothesis, statistical analysis has been performed in Orange, which is an open-source data mining toolbox for Python (Demsar *et al.* 2013). Linear regression allowed to investigate the relationship between selected variables, notably NAD Product and frequency per million (FreqMil). Due to the several outlying values in the FreqMil, the author applied logarithmic transformation (FreqLog) specifically, a natural log - before using linear regression (Figure 1).



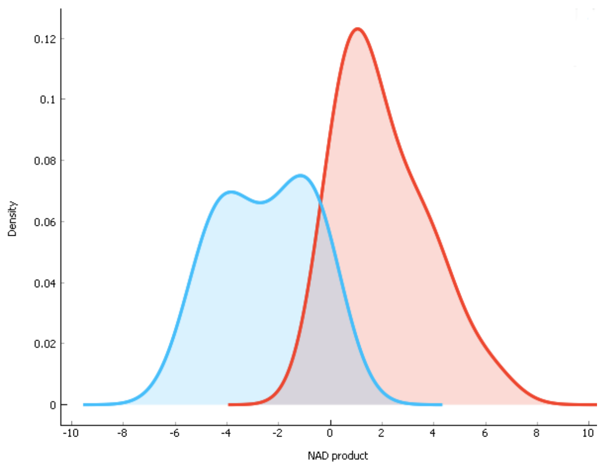
Distribution of FreqLog for double clusters (red = preferred; blue = dispreferred).



Distribution of FreqLog for triple clusters (red = preferred; blue = dispreferred).



Distribution of NAD Product for double clusters (red = preferred; blue = dispreferred).



Distribution of NAD Product for triple clusters (red = preferred; blue = dispreferred).

Figure 1. *Distribution FreqLog and NAD Product for double and triple clusters.*

For word-initial double clusters, the relationship between NAD Product and FreqLog is statistically significant only for the preferred clusters. Still, the correlation is low (r = 0.23). At the same time, there is no relationship between NAD Product and FreqLog for dispreferred clusters (r = 0). For triple clusters, the relationship between NAD Product and FreqLog is statistically significant for both, however the correlation for preferred and dispreferred consonant clusters is still low (r=0.26). The scatter plots are demonstrated in the Figure 2 Therefore, Hypothesis 2 has been validated only partially.

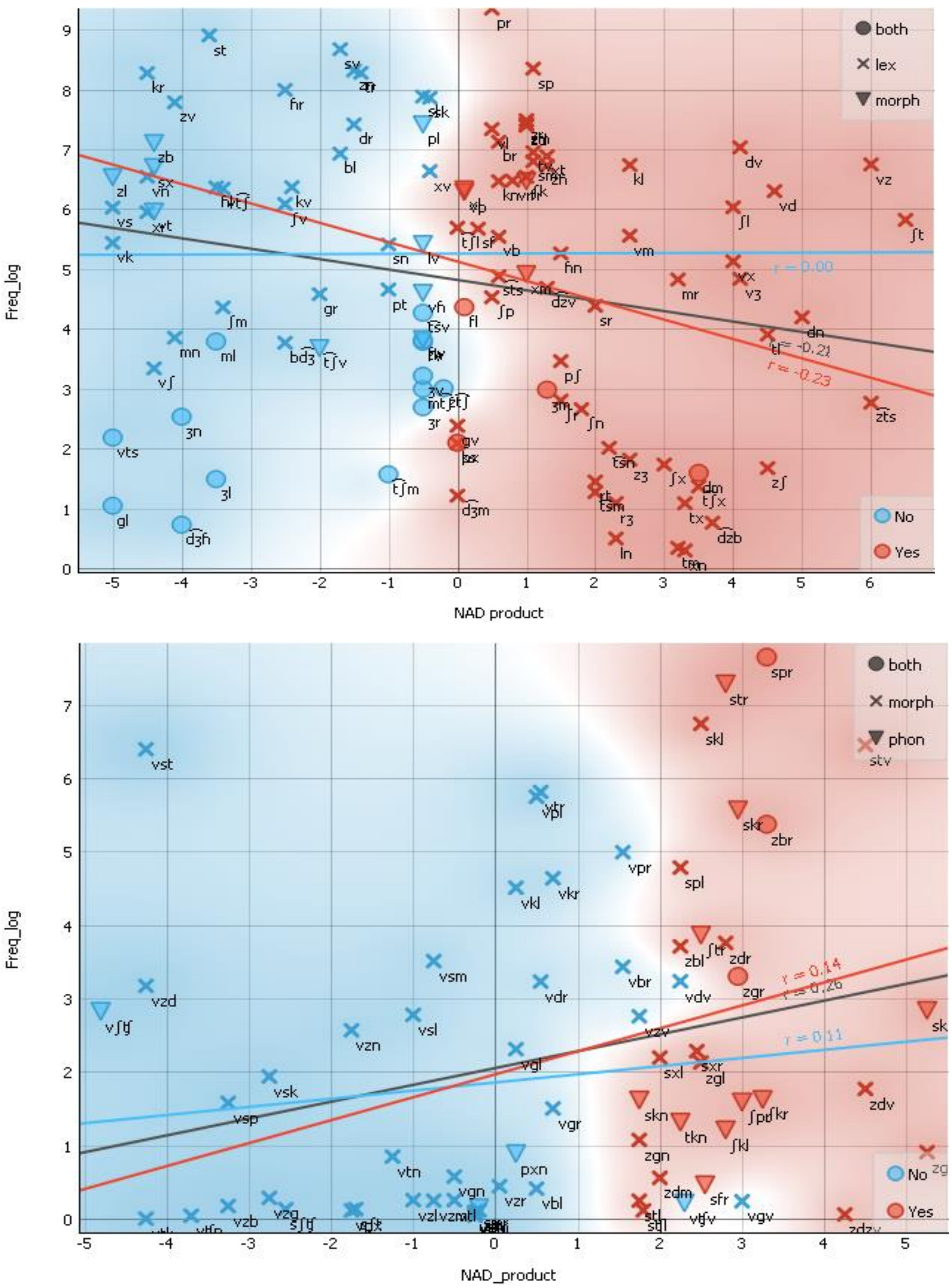


Figure 2. *Linear regression analysis for double and triple consonant clusters*

4. Conclusions

The general purpose of this pioneering research was to present, differentiate, and explain an overview of consonantal phonotactics of Ukrainian, contrasting it with morphonotactics. This is the first attempt to give a quantitative view of the state of morphological composition, preferability, and frequency of consonant clusters in the Ukrainian language. This corpus-based study relied on data from the huge electronic corpus GRAC, which allowed the author to provide the first quantitative generalizations about the distribution of morphological and lexical patterns of Ukrainian consonant clusters. Based on the quantitative analysis confirming a great inventory of consonant clusters, it can be concluded that Ukrainian is a consonantal language, but in the word-initial position there are fewer consonant clusters compared to Russian (Dressler & Kononenko-Szoszkiewicz, 2020) and Polish (Zydorowicz et. al. 2016). The main focus of the study was based, for the first time on, the phonological theory of Beats-and-Binding phonotactics developed by Dziubalska-Kołaczyk (2002), which allowed to include an analysis of the existence of consonant clusters. Two hypotheses were tested, which confirmed a general presumption that morphonotactic clusters tend to be marked and therefore dispreferred. Yet, the statistical analysis showed only a weak correlation between consonant clusters' frequency and their preference according to the NAD. The results of this study serve as a starting point for extending the research on Ukrainian morphonotactics in word-medial and word-final positions. The present study could be used as a foundation for comparative typological studies, research in the language acquisition, and processing of Ukrainian morphonotactic vs. phonotactic consonant clusters.

Bibliographical References

Andersen, Henning 1969. Indo-European voicing sandhi in Ukrainian. *Scando-Slavica* 15. 157–169.

Bilodid, Ivan 1969. *Sučasna ukrajinsjka literaturna mova. Vstup. Fonetyka* [Modern Ukrainian literary language. Introduction. Phonetics]. Kyiv: Naukova dumka.

Buk, Solomija, Jan Mačutek & Rovenchak, Andrij 2008. Some properties of the Ukrainian writing system. *Glottometrics* 16. 63–79.

Czaplicki, Bartlomiej 2007. Syllable Structure of Ukrainian. An OT Perspective. *Poznan Studies in Contemporary Linguistics* 2. 23–41.

Demsar Tomaz, Curk, Ales Erjavec, Crt Gorup, Tomaz Hocevar, Mitar Milutinovic, Martin Mozina, Matija Polajnar, Marko Toplak, Anze Staric, Miha Stajdohar, Lan Umek, Lan Zagar, Jure Zbontar, Marinka Zitnik & Blaz Zupan 2013. Orange: Data Mining Toolbox in Python, *Journal of Machine Learning Research* 14. 2349−2353.

Dressler, Wolfgang U. &. Dziubalska-Kołaczyk, Katarzyna 2006. Proposing morphonotactics. *Rivista di Linguistica* 18(2). 249-266.

Dressler, Wolfgang U. & Kononenko-Szoszkiewicz, Alona 2020. Main Differences Between German and Russian (Mor)phonotactics. A corpus-based study. In Wrembel, M., Kiełkiewicz-Janowiak, A., & Gąsiorowski, P. (eds.). *Approaches to the Study of Sound Structure and Speech: Interdisciplinary Work in Honour of Katarzyna Dziubalska-Kołaczyk.* Routledge.

Dressler, Wolfgang U., Hliničanová, Miroslava, Ďurčo, Matej, Mörth, Karlheinz & Korecky- Kröll, Katharina 2015. Phonotactic vs. morphonotactic obstruent clusters in Slovak and German. *Italian Journal of Linguistics* 27. 45–60.

Dziubalska-Kołaczyk, Katarzyna 2002. *Beats-and-Binding Phonology*. Frankfurt am Main: Peter Lang.

Dziubalska-Kołaczyk, Katarzyna & Krynicki, Grzegorz 2007. *Phonotactic Preferences in Polish, English and German: Quantitative Perspective.* (Paper presented at the 38th Poznan Linguistic Meeting, 13–16 Sept 2007).

Dziubalska-Kołaczyk, Katarzyna 2009. NP extension: B&B phonotactics. *Poznań Studies in Contemporary Linguistics* 45(1). 55–71.

Dziubalska-Kołaczyk, Katarzyna 2014. Explaining phonotactics using NAD. *Language Sciences* 46. 6–17.

Dziubalska-Kołaczyk, Katarzyna 2019. On the structure, survival and change of consonant clusters, *Folia Linguistica Historica* 53(40-1). 107–127.

Eckman, Fred. R. 1977. Markedness and the Contrastive Analysis Hypothesis. *Language Learning* 27. 315-330.

Isachenko, Aleksandr 1963. Opyt tipologicheskogo analiza slavjanskikh jazykov [Experience of typological analysis of Slavic languages]. *Novoe v lingvistike* 3. 106–121.

Kamandulytė-Merfeldienė, Laura 2015. Morphonotactics in L1 acquisition of Lithuanian: TD vs. SLI". *Eesti Rakenduslingvistika Ühingu aastaraamat* 11. 95-109.

Kelić Maja & Wolfgang U. Dressler 2019. The development of morphonotactic and phonotactic word–initial consonant clusters in Croatian first–language acquisition. *Suvremena lingvistika* 45. 179-200.

Klimenko, Nina, Karpilovska, Evgenia., Karpilovskyi, Viktor, & Nedozym, Tetiana 1998, Словник українських афіксальних морфів, *Dictionary of Ukrainian Affixal Morphemes,* Kyiv. O.O. Potebnia Institute of Linguistics of the National Academy of Sciences of Ukraine.

Korecky-Kröll Katharina, Dressler Wolfgang U., Freiberger Eva Maria, Reinisch Eva, Mörth Karlheinz & Gary Libben 2014. Morphonotactic and phonotactic processing in German-speaking adults. *Language Sciences* 46. 48-58.

Köpke, Barbara, Nocaudie, Olivier, Giraudo, Hélène & Basilio Calderone 2021. Exploring phonotactic and morphonotactic constraints in the acquisition of consonant clusters in L1 French. In Dressler, W. U., Calderone, B., Korecky-Kröll, K., & Sommer-Lolei, S. *Experimental and Acquisitional Approaches to Morphonotactics*, Austrian Academy of Sciences Press.

Ladefoged, Peter 2006. *A Course in Phonetics. 5th edition*. Boston: Heinle & Heinle.

Lewis, M. Paul, Simons, Gary F. & Fennig, Charles D. 2016. *Ethnologue: Languages of the world*, 19th edn. Dallas, SIL International

Majewicz, Alfred. F. 1989, *Języki świata i ich klasyfikaivanie*, Warszawa: Państwowe Wydawnictwo Naukowe.

Matvijas, I. H., I. O. Varcenko, P. Ju. Hrycenko, F. T. Zylko, Ja. V. Zak revs'ka, N. Jo. Marcuk, T. V. Nazarova, & N. P. Prylypko 1984. Atlas ukrajins'koji movy. 1: Polissja, Serednja Naddniprjanscyna i sumizni zemli. 1988. 2: Volyn', Naddnistrjanscyna i sumizni zemli. 2001. 3: Slobozanscyna, Doneccyna, Nyznja Naddniprjanscyna, Prycornomor"ja i sumizni zemli. Kiev: Naukova dumka.

Pompino-Marschall, Berndt, Steriopolo, Elena & Żygis, Marżena 2017. Ukrainian. Illustrations of the IPA*. Journal of the International Phonetic Association* 47. 349-357

Press, Igo & Pugh, Stefan 2015. *Ukrainian: A Comprehensive Grammar*. Taylor and Francis.

Prince, Alan & Smolensky, Prince 1993. *Optimality Theory: Constraint interaction in generative gramma*r. New Brunswick, NJ: Rutgers University Center for Cognitive Science

Sawicka, Irena 2001. *An outline of the phonetic typology of the Slavic languages*. Wydawnictwo Uniwersytetu M. Kopernika.

Shvedova, Maria, von Waldenfels, Ruprecht, Yarygin, Sergej, Rysin, Andriy, Starko, Vasyl & Nikolajenko, Timofij 2017-2022, *GRAC: General Regionally Annotated Corpus of Ukrainian. Electronic resource*, Kyiv, Lviv, Jena. Available at uacorpus.org.

Vakulenko, Maksym O. 2018. Ukrainian vowel system in the IPA context. *Govor* 35. 189– 214.

Vakulenko, Maksym O. 2019. Ukrainian Consonant Phones in the IPA Context with Special Reference to /v/ and /gh/. *Linguistica Online* 22. 36-61.

von Ahn, Louis 2022. Standing with Ukraine: a message from the CEO. *Duolingo blog*, Last access: 30. 05.2022, <https://blog.duolingo.com/duolingo-statement-ukraine/>.

Zhovtobrjukh, Mykhailo, A., & Kulyk, Borys, M. 1965. *Kurs suchasnoji ukrajinsjkoji literaturnoji movy [A course of moderm Ukrainian literary language]*. In 2 parts (3rd ed.). Kyjiv: Rad. shk. Part 1. (in Ukrainian).

Zilynsjkyj, Ivan 1932. *Opis fonetyczny języka ukraińskiego [Phonetic description of the Ukrainian language]*. Kraków: Gebethner and Wolff. (in Polish).

Zydorowicz, Paulina, Orzechowska, Paola, Jankowski, Michał, Dziubalska-Kołaczyk, Katarzyna, Wierzchoń, Piotr & Pietrala, Dawid 2016. *Phonotactics and morphonotactics of Polish and English: Theory, description, tools and applications*. Poznań: Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza.

Appendix

Table 5. Word-initial double consonant clusters

	Cluster	Lemma types	Tokens	Freq per mil	Ukrainian	Transliteration	English	Phon /morph/ both
1	pr	4923	10,000,000	11,616.16	право	pravo	right	phon
2	kr	2424	3,403,413	3,953.46	країна	krajina	country	phon
3	tr	2415	3,417,444	3,969.76	треба	treba	need	phon
4	sp	2325	3,671,492	4,264.86	спосіб	sposib	method	both
5	ɦr	1905	2,569,104	2,984.31	група	hrupa	group	phon
6	st	1812	6,391,604	7,424.59	стояти	stojaty	to stand	both
7	br	1504	1,079,423	1,253.88	брати	braty	to take	phon
8	dr	1423	1,441,753	1,674.76	другий	druhyj	second	phon
9	sl	1347	2,300,362	2,672.14	слово	slovo	word	phon
10	pl	1320	1,442,468	1,675.59	план	plan	plan	phon
11	sv	1296	5,071,125	5,890.7	свій	svij	your	phon
12	bl	1218	885,195	1,028.26	близько	blyz^{j}ko	near	phon
13	sk	1150	2,255,688	2,620.24	сказати	skazaty	to say	both
14	zn	1004	3,555,943	4,130.64	знати	znaty	to know	both
15	zv	1004	2,086,703	2,423.95	звичайно	zvyčajno	usually	both
16	dv	945	983,133	1,142.02	два	dva	two	phon
17	ɦl	841	503,06	584.36	глибокий	hlybokyj	deep	phon
18	kl	830	731,405	849.61	клас	klas	class	phon
19	xr	825	332,214	385.91	храм	xram	temple	phon
20	sm	804	778,427	904.23	смерть	smert'	death	phon
21	zm	804	1,456,451	1,691.84	зміна	zmina	change	both
22	zd	778	1,410,742	1,638.74	здаватися	zdavatysja	to seem	both
23	zl	704	596,521	692.93	злочин	zločyn	crime	phon
24	zb	620	1,054,493	1,224.92	збиратися	zbyratysja	gather	both
25	fr	615	559,743	650.21	фракція	frakcja	fraction	phon
26	vs	604	358,628	416.59	всякий	vsjakyj	any	both
27	kv	540	504,847	586.84	квітень	kviten'	april	phon
28	xl	520	486,246	564.83	хлопець	xlopec'	boy	phon
29	zr	504	1,548,212	1,798.43	зробити	zrobyty	to do	both
30	vn	497	602,235	699.97	внутрішній	vnutrišnij	internal	both
31	vr	495	564,634	655.89	враження	vraženja	impression	both
32	vl	484	1,334,128	1,549.74	влада	vlada	power	both
33	ʃt	471	291,924	339.1	штаб	štab	headquarters	phon
34	kn	440	558,386	648.63	книжка	knyžka	book	phon
35	zɦ	434	723,892	840.88	згадати	zxadaty	to remind	morph
36	ln	428	428	0.66	льняний	l'njany	linen	phon
37	vp	427	465,117	540.29	вперше	vperše	for the first time	both
38	sx	417	706,119	820.24	схожий	sxožy	similar	both
39	vz	417	739,380	858.88	взагалі	vzagali	in general	both
40	vt	405	335,139	389.3	втім	vtim	however	morph
41	ʃk	404	601,292	698.47	школа	škola	school	phon
42	tv	349	904,477	1054.14	твій	tvij	yours	phon

43	xv	344	660,630	767.4	хвилина	xvylyna	minute	phon
44	vd	330	472,838	549.26	вдатися	vdatysja	to suceed	both
45	ʃp	325	79,533	92.39	шпиталь	špytal'	hospital	phon
46	vt͡ʃ	323	492,694	572.32	вчений	včenyj	scientist	both
47	ɦn	315	166,349	193.23	гнів	hniv	anger	phon
48	fl	299	67,341	78.22	флот	flot	fleet	phon
49	mn	297	40,076	46.55	множина	množyna	plural	phon
50	vm	294	224,019	260.22	вміти	vmity	to be able to do	morph
51	sn	287	193,390	224.64	сніг	snix	snow	phon
52	sr	275	68,857	79.99	срібло	sriblo	silver	phon
53	gr	251	803,809	97.35	ґрунт	grunt	soil	phon
54	ʃv	250	381,341	442.97	швидко	švydko	fast	phon
55	ʃl	241	360,783	419.09	шлях	šljax	way	phon
56	vk	215	198,688	230.8	вказувати	vkazuvaty	to point	morph
57	t͡sv	172	61,178	71.07	цвинтар	cvyntar	cemetery	phon
58	xm	153	117,594	136.6	хмара	xmara	cloud	phon
59	ml	149	37,552	43.62	млин	mlyn	mill	phon
60	vb	133	219,907	255.45	вбити	vbyty	to kill	morph
61	pt	131	90,636	105.28	птах	ptax	bird	phon
62	ʃm	130	66,990	77.82	шматка	šmatka	piece	phon
63	vɦ	124	85,48	99.29	вгору	vgoru	uphill	phon
64	mr	122	107,361	124.71	мрія	mrija	dream	phon
65	dn	112	56,692	65.85	днями	dnjamy	days	phon
66	ɦv	111	38,957	45.25	гвардія	hvardija	guard	phon
67	sf	110	251,815	292.51	сфера	sfera	sphere	phon
68	zt͡ʃ	96	16,706	19.41	зчинитися	sčynytysja	to appear	morph
69	s͡ts	83	113,479	131.82	сцена	scena	stage	phon
70	d͡zv	82	93,229	108.3	дзвонити	dzvonyty	to call	phon
71	ks	81	6,214	7.22	ксенофобія	ksenofobija	xenophobia	phon
71	vʒ	77	198,529	126.07	вживати	vžyvaty	to use	morph
73	t͡ʃv	76	33,512	38.93	чверть	čwert'	quarter	phon
74	vʃ	75	23,642	27.47	вшанувати	všanuvaty	to honor	morph
75	bd͡ʒ	70	36,728	42.66	бджола	bdžolɑ	bee	phon
76	xt	68	830,482	984.7	хтось	xtos'	someone	phon
77	pʃ	66	26,824	31.16	пшениця	pšenycja	wheat	phon
78	ʒm	64	16,304	18.94	жменю	žmeniu	a handful	phon
79	vx	60	145,509	169.03	входити	vxodyty	to enter	morph
80	rv	58	37,355	43.39	рватися	rvatysja	to tear	phon
81	z͡ts	57	12,943	15.03	зцілення	zcilennja	healing	morph
82	tl	56	42,274	49.11	тлумачення	tlumačennja	translation	phon
83	tk	53	38,508	44.73	тканини	tkanyny	fabrics	phon
84	ʃn	49	11,533	13.4	шнур	šnur	cord	phon
85	zʒ	47	4,479	5.2	зжерти	zžerty	devour	morph
86	ʃr	45	11,383	15.62	шрам	šram	scar	phon
87	gl	41	1,611	1.870	ґлорія	glorija	glory	phon
88	gv	41	8,521	9.9	ґвалт	gvalt	uproar	phon

89	t͡ʃm	41	3,316	3.85	чмихнути	čmyxnuty	to snicker	phon
90	ʒr	35	11,953	13.88	жрець	žrec'	votary	phon
91	rʒ	33	1726	2	ржати	ržaty	to growl	phon
92	t͡ʃl	28	256,099	297.49	член	člen	member	phon
93	ʒv	27	20,829	24.20	жвавий	žvavyj	alive	phon
94	zʃ	20	3,771	4.38	зшиток	sšytok	notebook	morph
95	ʒl	20	2,992	3.480	жлоб	žlob	parasite	phon
96	t͡sm	19	2,236	2.6	цмокнути	cmoknuty	to smack	phon
97	dm	17	3,416	3.970	дмухнути	dmuxnuty	to blow	phon
98	px	17	6,150	7.14	пхати	pxaty	push	phon
99	vts	17	6,854	7.96	вціліти	vcilyty	to survive	morph
100	lv	16	147,519	226.11	львівський	l'vivskyj	from Lviv	phon
101	tm	15	365	0.42	тмин	tmyn	cumin	phon
102	ʃx	14	4,055	4.71	шхуна	šxuna	schooner	phon
103	ʒn	12	9,534	11.7	жнива	žnyva	harvest	phon
104	t͡sn	11	5,638	6.55	цнота	cnota	virtue	phon
105	tx	9	1,712	1.99	тхір	txir	ferret	phon
106	d͡zb	7	1000	1.16	дзбан	dzban	pitcher	phon
107	mt͡ʃ	7	16,552	19.19	мчати	mčaty	race	phon
108	d͡ʒɦ	6	940	1.09	джгут	džxut	plait	phon
109	d͡ʒɦ	6	2,060	2.39	джміль	džmil'	bumblebee	phon
110	rt	6	2,842	3.3	ртуть	rtut'	mercury	phon
111	t͡ʃx	4	2,552	2.96	чхати	čxaty	to sneeze	phon
112	xn	3	307	0.36	хникати	xnykaty	to weep	phon

Table 6. Word-initial triple consonant clusters

	Cluster	Types	Tokens	Freq per mil	Ukrainian	Transliteration	English	Phon /morph/ both
1	str	1672	1,265,194	1469.67	структура	struktura	structure	phon
2	spr	563	1,823,217	2117.88	справа	s+prawa	right	both
3	skr	450	2,262,82	262.85	скрізь	skriz'	through	phon
4	skl	351	733,725	852.31	склад	s+klad	warehouse	morph
5	spl	214	102,828	119.45	сплачувати	s+plačuvaty	to pay	morph
6	vst	160	519,144	603.05	встановити	v+stanovyty	to set	morph
7	zdr	134	36,349	42.22	здригнутися	z+drygnutusia	to shudder	morph
8	zbr	107	186,354	216.47	зброя	zbroja	weapon	both
9	vpr	104	126,992	147.52	впродовж	v+pro+dovž	during	morph
10	vtr	89	289,774	336.61	втратити	v+tratyty	to loose	morph
11	stv	77	548,457	637.1	створення	s+tvorenia	creation	morph
12	vkr	71	88,857	103.22	вкрай	v+krai	extremely	morph
13	ʃtr	70	40,511	47.06	штраф	štraf	fine	phon
14	zɦr	66	22,717	26.39	зграя	zgraja	flock	both
15	skv	64	13,975	16.23	сквер	skver	square	phon
16	zbl	64	34,599	40.19	зближення	z+blyženia	rapprochement	morph
17	vkl	63	78,013	90.62	вкладати	v+kladaty	to invest	morph

18	vpl	62	271,526	315.41	вплив	v+plyv	influence	morph
19	vsp	56	3,347	3.89	вспіти	v+spity	to be on time	morph
20	zɦl	44	6,432	7.47	зглянутися	z+glianutysia	to take a look	morph
21	vɦl	37	7,882	9.16	вглиб	v+glyb	deeply	morph
22	sxr	34	7,655	8.89	схрестити	s+hrestyty	to cross	morph
23	ʃkr	33	3,543	4.12	шкребти	škrebty	to scratch	phon
24	vzd	31	19,840	23.05	вздовж	v+z+dovž	along	morph
25	vsl	30	13,090	15.21	вслід	v+slid	followed by	morph
26	vsm	30	28,171	32.72	всміхнутися	v+smixnutysia	to smile	morph
27	zdv	27	4,230	4.91	здвигнути	z+dvygnuty	to move	morph
28	vɦr	25	3,039	3.53	вгризатися	v+gryzatysia	to gnaw into	morph
29	vʃʧ	22	13,764	15.99	вщент	vščent	to smash to atoms	phon
30	vsk	20	5,162	6	вскочити	v+skočyty	to jump in	morph
31	vzr	20	490	0.57	взріти	v+zrity	to notice	morph
32	vdv	19	21,188	24.61	вдвічі	v+dviči	twice	morph
33	vzn	18	10,454	12.14	(давати) взнаки	v+znaky	to show up	morph
34	vdr	17	21,076	24.48	вдруге	v+druge	a second time	morph
35	zɦn	17	1,679	1.95	згнити	z+gnyty	to rotten	morph
36	skn	16	3,509	4.08	Скнара	sknara	miser	phon
37	sxl	15	6,941	8.06	схлипувати	s+hlypuvaty	to sob	morph
38	vɦn	15	679	0.79	вгніздитися	b+gnizdytysia	to nest	morph
39	vbr	13	25,970	30.17	вбрання	v+brania	cloth	morph
40	vzv	10	12,826	14.9	взвод	v+z+vod	platoon	morph
41	stl	9	246	0.29	стлумити	s+tlumyty	to opress	morph
42	ʃpr	8	3,395	3.94	шприц	špritz	syrenge	phon
43	zɦv	8	1,293	1.5	згвалтувати	z+gvaltuvaty	to rape	morph
44	ʃkl	7	2,050	2.38	шклянка	šklanka	glass	phon
45	smr	7	167	0.17	смрад	smrad	stench	phon
46	vɦv	7	237	0.28	вгвинчуватися	v+gvynčuvatysia	to screw	morph
47	sfr	6	514	0.6	сфрагістика	sfragistyka	sphragistics	phon
48	sʃt	6	127	0.15	зштовхнути	z+štovhnuty	to push away	morph
49	vzl	6	254	0.3	взлісся	v+z+lisia	outskirt	morph
50	tkn	5	2,380	2.76	ткнути	tknuty	to poke	phon
51	vzb	5	176	0.2	взбіччі	v+z+biči	on the sidelines	morph
52	vzg	5	296	0.34	взгір'я	v+z+girja	hill	morph
53	zdm	5	652	0.76	здмухнути	z+dmuxnuty	to blow away	morph
54	zmr	5	115	0.13	змружити	z+mružyty	screw up ones eyes	morph
55	pxn	4	1,244	1.45	пхнути	pxnuty	to push	phon
56	sʃʧ	4	132	0.15	зщулився	s+ščulyty	to shrink	morph
57	vbl	4	445	0.52	вблагати	v+blagaty	to beg	morph
58	vʧp	4	47	0.05	вшпарити	v+šparyty	to do sth energetically	morph
59	vʧv	4	211	0.25	вчвал	včval	galloping	phon
60	sʧl	3	136	0.13	зчленування	s+členuwania	jointing	morph
61	vdm	3	92	0.11	вдмухнути	v+dmuhnuty	to blow	morph
62	vpn	3	47	0.05	впнути	v+pnuty	to stick	morph

63	vpx	3	294	0.13	впхати	v+pxaty	to squeeze in	morph
64	vtl	3	258	0.3	втлумачити	v+tlumačyty	to interpret	morph
65	vzm	3	247	0.29	взмозі	v+z+mozi	able to	morph
66	vtn	2	1,161	1.35	втнути	v+tnuty	to cut out	morph
67	vxl	2	69	0.08	вхлинав	v+xlynav	to consume	morph
68	zʣv	2	56	0.07	здзвонитися	z+dzwonytysia	to call	morph
69	vtk	1	10	0.01	вткати	v+tkaty	to stick	morph