

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

WOLFGANG U. DRESSLER^{1,2}
ALONA KONONENKO-SZOSZKIEWICZ¹

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 Zygorowicz 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 /pf/ (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 and 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:

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

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 calcula-

tor in order to assign a preferability index which is “a number denoting a degree to which a given preference is observed” (Dziubalska-Kołączyk 2019). The formula for word-initial consonant clusters is as follows:

$$\text{NAD product} = \text{NAD C1C2} - \text{NAD C2V}$$

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. *Kunst* ‘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, Mörtz & 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/),

/-mpfst/ (10): *kämpf-st* ‘(you) fight’,

/-ŋfst/ (3): *wünsch-st* ‘(you) wish’,

/-ntfst/ (3): *plantsch-st* ‘(you) splash’, recent English loan words *launch-st*, *lunch-st*. In oral speech, the /s/ is most often reduced after /ʃ, tʃ/ 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’,

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

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

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

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

/-l̥tsst/ (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:

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

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

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

/-rnsts/ (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 *stülپ-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	Vr <u>p</u> st	204	1,095,735	0.02
2	Vr <u>f</u> st	65	11,421	0.57
3	Vm <u>p</u> st	11	1,101	1
4	V <u>ŋ</u> kst	37	9,688	0.38
5	Vr <u>k</u> st	30	5,149	11.38
6	Vr <u>m</u> st	29	255	10.38
7	Vr <u>x</u> st	11	106	1.29
8	Vm <u>p</u> fst	10	776	10.39
9	Vl <u>m</u> st	8	77	0.95
10	Vr <u>n</u> st	8	< 1,200,000	
11	Vl <u>k</u> st	5	526	0.82
12	Vn <u>f</u> st	5	607	0.73
13	Vl <u>p</u> st	5	687	0.73
14	Vl <u>f</u> st	4	828	0.48
15	Vx <u>t</u> sst	3	3	100
16	Vl <u>x</u> st	2	2	100
17	Vl <u>l</u> fst	2	9	22.22
18	Vf <u>t</u> sst	2	9	22.22
19	Vr <u>t</u> sst	2	4	50
20	Vn <u>t</u> fst	1	1	100
21	Vl <u>t</u> sst	1	1	100
22	V <u>ŋ</u> ksts	1	23	4.35
23	Vr <u>p</u> sts	1	1,835	0.05
24	Vl <u>p</u> sts	1	25	4
25	Vr <u>n</u> sts	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 (McEnery & 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 excep-

tions, there is a group of clusters /-lkst, -ŋfst, -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, -mpfst/, 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 /ts/. Due to such fusion,

genitives ending in /ts/ 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’,

/-xtst/: 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’,

/-fst/: *wisch-st* ‘(you) wipe’,

/-pfst/: *klopf-st* ‘(you) knock’,

/-tjst/: *rutsch-st* ‘(you) slip’,

/-ftst/: 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’

/-ltst/: *walz-t* ‘(s)he waltzes’ and its participle,

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

/-lft/: only in *fälsch-t* ‘(s)he falsifies’ and its participle and derived verbs,

/-mft/: only in *ramsche-t* ‘(s)he buys cheap junk’ and its participle and derived verbs,

/-rtft/: only in *turtsch-t* ‘taps (eggs)’ and its participle,

/-njft/: *wünsch-t* ‘(s)he wishes’ and its participle,

/-pft/: *grapsch-t* ‘grabs’ and its past participle,

/-rft/: *forsch-t* ‘(s)he researches’ and its participle,

/-ntft/: *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,

/-mpft/: *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.

/-rtst/ 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 participle *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	ηkt	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	rtst	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	ntst	266	560,076	0.05
11	xst	246	1,838,731	0.01
12	mpft	232	662,652	0.03
13	mst	226	164,703	0.14
14	lkt	182	2,809,304	0.01
15	rft	163	625,920	0.03
16	ltst	156	54,562	0.29
17	fst	136	89,308	0.15
18	rkt	134	1,358,674	0.01
19	rxst	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	ɲft	45	354,583	0.01
25	fst	40	1,388	2.89
26	tʃst	31	150	20.7
27	pft	27	5075	0.41
28	fst	23	19,692	0.12
29	lft	22	52,109	0.04
30	xtst	19	9,353	0.2
31	ntʃt	19	9,188	0.21
32	lxt	16	1,580	1.01
33	mft	10	770	1.3
34	nxt	5	2066	0.24
35	rtʃt	4	4	100
36	pfst	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 has 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 NAD³

Nº	IPA transcription	NAD (VC)	NAD (C1C2)	NAD (C2C3)	NAD product	Preferred cluster?
1	Vrpt	2	6.6	1	5.1	Yes
2	Vrtst	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 /-ntʃt/, /-tʃst/ and /rtʃt/ were excluded from the analysis because the NAD calculator does not recognize affricate /-tʃ/. 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	Vrjt	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	Vltst	2.5	3	0.5	1.5	Yes
13	Vrxt	2	4.4	4	1.4	Yes
14	Vmjt	3	3.5	1.5	1.25	Yes
15	Vmpft	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	Vljt	2.5	3	1.5	1	Yes
20	Vxtst	5	3.5	0.5	0.75	No
21	Vntst	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	Vnjt	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	Vpjt	6	2.5	1.5	-1.25	No
29	Vpst	6	2	1	-1.5	No
30	Vftst	5	1	0.5	-1.75	No
31	Vpfst	5.5	1	1	-2.25	No
32	Vfst	5	0.5	1	-2.5	No
33	Vjst	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 /rtst/ 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

Variables	Factor loadings (Varimax normalized) Extraction : Principal components (Marked loadings are > .700000)		
	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,

/-rfs : *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',

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

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

/-ntfs/: 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’,

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

/-mjfs/ 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	mpfs	20	35,000	0.56%
8	rxs	7	70,000	0.01%
9	rjs	7	3,300	0.21%
10	lxs	6	268	2.24%
11	ntfs	5	370	1.35%
12	sks	5	145	3.45%
13	njs	2	176	1.14%
14	mjs	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
Vjks	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 */ts/*, due to Gen.SG. and rarely PL *-s*: */xts/* as in *Berichts* ‘report’, */kts/* as in *Projekts* ‘project’, */pts/* as in *Konzepts* ‘concept’, */lts/* as in *Anwalts* ‘lawyer’, */nts/* as in *Abends* ‘in the evening’, and */rts/* 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 *-TS*

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):

/-rst̪s/: *Durst-s* ‘thirst’,

/-lst̪s/: *Schwulst-s* ‘bombast’,

/-pst̪s/: *Papst-s* ‘pope’, *Herbst-s* ‘autumn’ and their many compounds,

/-nst̪s/: *Dienst-s* ‘service’ and its many compounds,

/-rk̪ts/: *Markt-s* ‘market’ and its many compounds,

/-nk̪ts/: *Punkt-s* ‘point’ and its many compounds,

/-nft̪s/: *Senft-s* ‘mustard’,

/-rpt̪s/ only in *Exzerpt-s* ‘excerpt’,

/-tst̪s/ only in *Arzt-s* ‘physician’ with its many compounds,

/-kst̪s/ only in *Text-s* and its compounds

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

№	Cluster	Types	Tokens	TTR
1	<u>rsts</u>	8	2,434	0.33%
2	psts	5	3,085	0.16%
3	<u>lsts</u>	5	389	1.29%
4	<u>nsts</u>	4	8,800	0.04%
5	<u>rkts</u>	2	58,097	0.003%
6	nfs	2	301	0.66%
7	<u>ksts</u>	1	2,000	0.05%
8	<u>rtsts</u>	1	407	0.24%
9	<u>rpts</u>	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 morpho-notactic.

2.5. WORD-INITIAL POSITION

The German standard has no monoconsonantal prefixes, in contrast to Bavarian-Austrian dialects, as in *g'storben* 'died', *b'soffen* '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-Kolaczyk (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’, /sts-/ 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 /tʃv-/ 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 pfl as in *pflegen* ‘to care for’, /pfr-/ 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 /tʃv-/ 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-Kořaczyk 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, pfl, pfr, fl fʋ, fr, fm, fn, ps, sf, sm, sts, tsw/ 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 /spr/, which has the highest token frequency and the second-highest type frequency; the next cluster in the hierarchy of preferences is /str/, 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 /str/ and /spr/ 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 /spr/ 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, rtsn, 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 *herbstlich* ‘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-ts, 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 monosyllabic or bisyllabic. 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*, *-ts*. 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. 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 Zygorowicz 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, ts/, 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-Kolaczyk, 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-Kolaczyk, Katarzyna (2006) Proposing morphonotactics, *Italian Journal of Linguistics* 18, 249–266.
- Dressler, Wolfgang U.; Dziubalska-Kolaczyk, Katarzyna & Pestal, Lina (2010) Change and Variation in Morphonotactics, *Folia Linguistica Historica* 31, 51–67.
- Dressler, Wolfgang U.; Hliničanová, Miroslava; Ďurčo, Matej; Mörtz, Karlheinz & Korecky-Kröll, Katharina (2015) Phonotactic vs. morphonotactic obstruent clusters in Slovak and German, *Italian Journal of Linguistics* 27, 45–60.
- Dressler, Wolfgang U.; Ketz, 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-Kolaczyk, Katarzyna (2002) *Beats-and-Binding Phonology*. Frankfurt: Lang.
- Dziubalska-Kolaczyk, Katarzyna (2009) NP extension: B&B phonotactics, *PSiCL* 45, 55–71.
- Dziubalska-Kolaczyk, Katarzyna (2014) Explaining phonotactics using NAD, *Language Sciences* 46, 6–17.
- Dziubalska-Kolaczyk, Katarzyna (2019) On the structure, survival and change of consonant clusters, *Folia Linguistica Historica* 53(s40-1), 107–127.
- Dziubalska-Kolaczyk, 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-Kolaczyk, Katarzyna & Weckwerth, Jarosław (2002) *Future Challenges for Natural Linguistics*. Munich: Lincom.
- Fehrer, 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 français (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; Mörrth, 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 congrès 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; Mörrth, 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 extra-syllabic 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; Wierchoń, Piotr & Pietrala, Dawid (2016) *Phonotactics and morphonotactics of Polish and English. Theory, description, tools and applications*. Poznań: Adam Mickiewicz University Press.