DOI: 10.26346/1120-2726-219

(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 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highlighted, the central issue of Ukrainian phonetics is that contemporary judgments about the language are based on outdated phonetic material obtained from just one speaker and processed with old-fashioned phonetic methods.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 the 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.

Another 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. Zilynski (1979) 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 presents 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 interpreted with regard to 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).

The syllable structure of Ukrainian has been analyzed by Czaplicki (2007) from the Optimality Theory perspective (Prince & 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 the genitive 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-and-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 consonant clusters.

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-*, *ekstra--*, *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* (**GRAC** – in Ukr.: 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/.

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 – 81 – of clusters are phonotactic, and 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,

IPA	Ukr	b	ʋ	ɦ	g	d	ʒ	ʣ	ʤ	z	k	l	m	n	p	r	s	t	f	x	ts	ʧ	ʃ
		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	-	+	-	-	-	-	-	-	-	-	+	+	+	-	+	-	-	-	-	-	-	-
ʣ	дз	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ʤ	дж	-	-	+	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
z	з	+	+	+	-	+	+	-	-	-	-	+	+	+	-	+	-	-	-	-	+	+	+
k	к	-	+	-	-	-	-	-	-	-	-	+	-	+	-	+	+	-	-	-	-	-	-
l	л	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-
m	м	-	-	-	-	-	-	-	-	-	-	+	-	+	-	+	-	-	-	-	-	+	-
n	н	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
p	п	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	+	-	+	-	-	+
r	р	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
s	с	-	+	-	-	-	-	-	-	-	+	+	+	+	+	+	-	+	+	+	+	-	-
t	т	-	+	-	-	-	-	-	-	-	+	+	+	-	-	+	-	-	-	+	-	-	-
f	ф	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-
x	х	-	+	-	-	-	-	-	-	-	-	+	+	+	-	+	-	+	-	-	-	-	-
ts	ц	-	+	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-
ʧ	ч	-	+	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	+	-	-	-
ʃ	ш	-	+	-	-	-	-	-	-	-	+	+	+	+	+	+	-	+	-	+	-	-	-

Table 1. Combinatory possibilities of Ukrainian phonemes.

namely /sp-/, /st-/, /sk-/, /zn-/, /zv-/, /zm-/, /zd-/, /zb-/, /vs-/, /zr-/, /vn-/, /vr-/, /vl-/, /vp-/, /sx-/, /vz-/, /vd-/, /vʧ-/. 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 (Appendix, Table 6). Only three consonant clusters, namely /spr-/, /zbr-/, /zfir-/ 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 + fir/ in *z + gribaty* 'to shovel, /zfir/ *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 something 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

	IPA	b	ʋ	ɦ	g	d	ʒ	ʣ	ʤ	z	k	l	m	n	p	r	s	t	f	x	ʦ	ʧ	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	-	-	-	-	-	-	-
ʣ	дз	no	yes	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ʤ	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	-	-	-	-	-
ʦ	ц	-	yes	-	-	-	-	-	-	-	-	-	yes	no	-	-	-	-	-	-	-	-	-
ʧ	ч	-	yes	-	-	-	-	-	-	-	-	yes	yes	-	-	-	-	-	-	no	-	-	-
ʃ	ш	-	yes	-	-	-	-	-	-	-	yes	no	yes	no	no	yes	-	no	-	no	-	-	-

Table 2. NAD preferences for word-initial doubles.

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. Hypothesis 2 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 Table 2, among 112 word-initial double consonant clusters, 61 clusters are preferred, and 51 are dispreferred.

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 (ʊk-, ʊʒ-, ʊʃ-, ʊx-)	N = 9 (zh-, ʊt-, ʊm-, ʊb-, ztʃ-, zts-, zʒ-, zʃ-, ʊts-)
Phon	N = 53 (pr-, kr-, tr-, hr-, br-, dr-, sl-, pl-, sʊ-, bl-, dʊ-, hl-, kl-, xr-, sm-, zl-, fr-, kʊ-, xl-, kn-, tʊ-, xʊ-, hn-, fl-, sr-, gr-, ʃʊ-, ʃl-, tsʊ-, xm-, ʃm-, ʊh-, mr-, dn-, hʊ-, ʤʊ-, tʃʊ-, ʒm-, rʊ-, tl-, ʃr-, gl-, gʊ-, tʃm-, ʒr-, tʃl-, ʒʊ-, ʒl-, tsm-, dm-, lʊ-, tm-, xn-)	N = 28 (ʃt-, ln-, ʃk-, ʃp-, mn-, sn-, ml-, pt-, sf-, sts-, ks-, bʤ-, xt-, pʃ-, tk-, ʃn-, rʒ-, px-, ʃx-, ʒn-, tsn-, tx-, ʤb-, mtʃ-, ʤh-, ʤh-, rt-, tʃx-)
Both	N = 4 (zʊ-, zm-, zr-, ʊtʃ-)	N = 14 (sp-, st-, sk-, zn-, zd-, zb-, ʊs-, ʊn-, ʊr-, ʊl-, ʊp-, sx-, ʊz-, ʊd-)

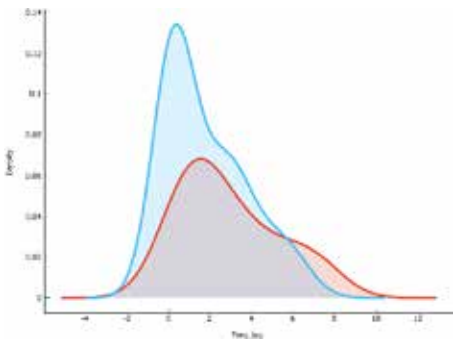
Table 3. NAD preferences of word-initial doubles.

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). However, the five most common consonant combinations are of a morphonotactic type and preferred according to the NAD. (Table 4).

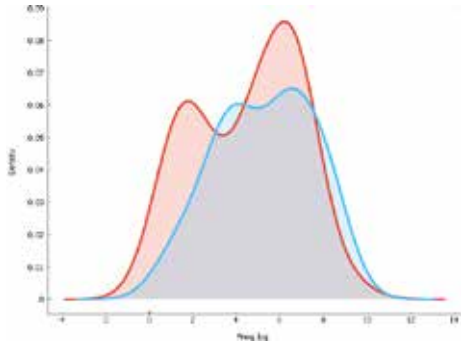
	PREFERRED	DISPREFERRED
Morph	N = 15 (skl-, spl-, zdr-, stʊ-, zbl-, zxl-, sxr-, zdʊ-, zxn-, shl-, stl-, zxʊ-, zdm-, stʃl-, zdʒʊ-)	N = 37 (ʊst-, ʊpr-, ʊtr-, ʊkr-, ʊkl-, ʊpl-, ʊsp-, ʊxl-, ʊzd-, ʊsl-, ʊsm-, ʊxr-, ʊsk-, ʊzr-, ʊdʊ-, ʊzn-, ʊdr-, ʊxn-, ʊbr-, ʊzʊ-, ʊxʊ-, sʃt-, ʊzl-, ʊzb-, ʊzx-, zmr-, sʃtʃ-, ʊbl-,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

Table 4. NAD preferences for word-initial triples.

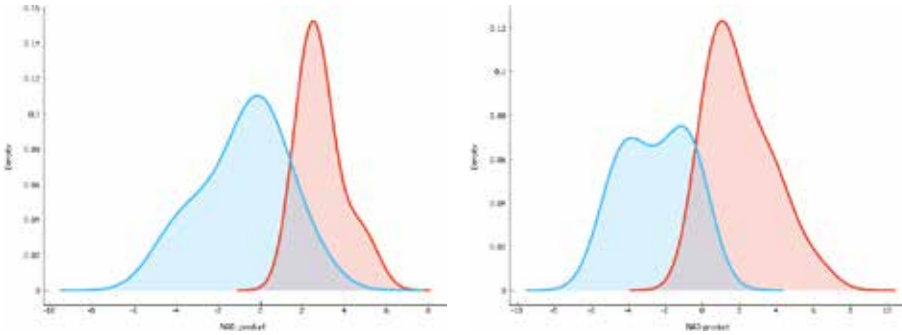
To validate Hypothesis 2, statistical analysis has been performed in Orange, which is an open-source data mining toolbox for Python (Demsar *et al.* 2013). Linear regression allowed to investigate the relationship between selected variables, notably NAD Product and frequency per million (FreqMil). Due to the several outlying values in the FreqMil, a logarithmic transformation (FreqLog), specifically the natural log, was applied before performing linear regression. (Figure 1).



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



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

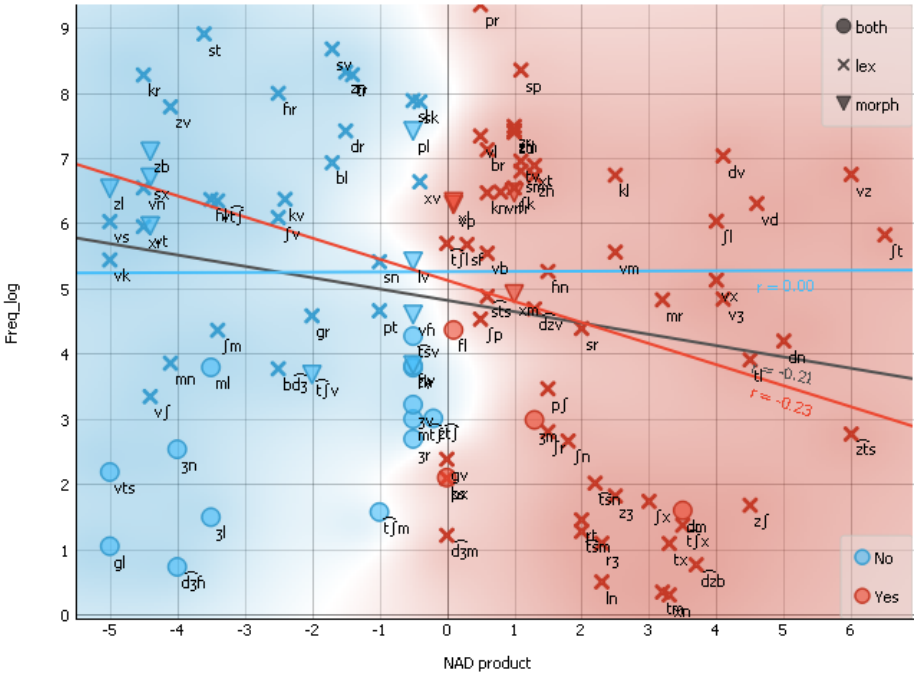


Distribution of NAD Product for double clusters (red = preferred; blue = dispreferred).

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



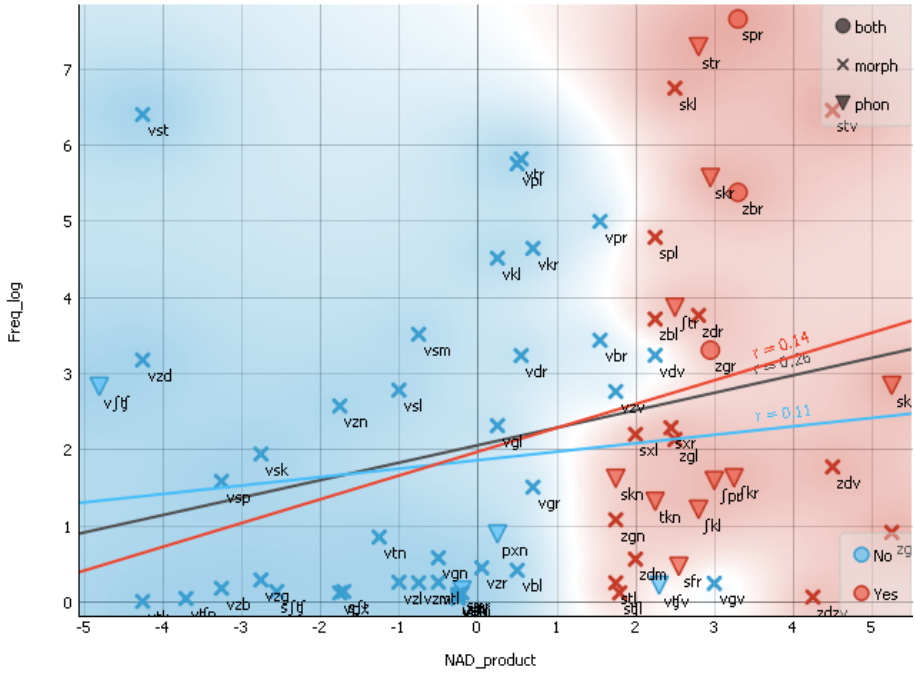


Figure 2. Linear regression analysis for double and triple consonant clusters.

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.

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; Mačutek, Jan & Rovenchak, Andrij 2008. Some properties of the Ukrainian writing system. *Glottometrics* 16. 63-79.

Czaplicki, Bartlomiej 2007. Syllable Structure of Ukrainian. An OT Perspective. *Poznań Studies in Contemporary Linguistics* 2. 23-41.

Demsar Tomaz, Curk; Erjavec, Ales; Gorup, Crt; Hocevar, Tomaz; Milutinovic, Mitar; Mozina, Martin; Polajnar, Matija; Toplak, Marko; Staric, Anze; Stajdohar, Miha; Umek, Lan; Zagar, Lan; Zbontar, Jure; Zitnik, Marinka & Zupan, Blaz 2013. Orange: Data Mining Toolbox in Python. *Journal of Machine Learning Research* 14. 2349-2353.

Dressler, Wolfgang U. & Dziubalska-Kołaczyk, Katarzyna 2006. Proposing morphonotactics. *Italian Journal of Linguistics* 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 September 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 & Dressler, Wolfgang U. 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; Karlheinz, Mörth & Libben, Gary 2014. Morphonotactic and phonotactic processing in German-speaking adults. *Language Sciences* 46. 48-58.

Köpke, Barbara; Nocaudie, Olivier; Giraudo, Hélène & Calderone, Basilio 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 ed. Dallas: SIL International.

Majewicz, Alfred. F. 1989, *Języki świata i ich klasyfikaivanie*. Warszawa: Państwowe Wydawnictwo Naukowe.

Matvijas, I. H.; Varcenko, I. O.; Hrycenko, P. Ju.; Zylko, F. T.; Zak revs'ka Ja. V.; Marcuk, N. Jo.; Nazarova, T. V. & Prylypko, N. P. 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 grammar*. 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*. <blog.duolingo.com/duolingo-statement-ukraine>. Last access: 30.05.2022.

Zhovtobrjukh, Mykhailo A. & Kulyk, Borys M. 1965. *Kurs suchasnoji ukrajinsjkoji literaturnoji movy* [A course of modern Ukrainian literary language]. In two parts (3rd ed.). Kyjiv: Rad. shk. Part 1.

Zilynsjkyj, Ivan 1932. *Opis fonetyczny języka ukraińskiego* [Phonetic description of the Ukrainian language]. Kraków: Gebethner and Wolff.

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	TRANSLATION	PHON /MORPH
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	fr	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	blyzʹko	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	fl	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	smertʹ	death	phon

	CLUSTER	LEMMA TYPES	TOKENS	FREQ PER MIL	UKRAINIAN	TRANSLITERATION	TRANSLATION	PHON /MORPH
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	zfi	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

	CLUSTER	LEMMA TYPES	TOKENS	FREQ PER MIL	UKRAINIAN	TRANSLITERATION	TRANSLATION	PHON /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	vtʃ	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	грунт	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	tsv	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

	CLUSTER	LEMMA TYPES	TOKENS	FREQ PER MIL	UKRAINIAN	TRANSLITERATION	TRANSLATION	PHON /MORPH
62	ʃm	130	66,990	77.82	шматка	šmatka	piece	phon
63	vfi	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	fiv	111	38,957	45.25	гвардія	hvardija	guard	phon
67	sf	110	251,815	292.51	сфера	sfera	sphere	phon
68	ztʃ	96	16,706	19.41	зчинитися	sčynytysja	to appear	morph
69	sts	83	113,479	131.82	сцена	scena	stage	phon
70	dzv	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	tʃv	76	33,512	38.93	чверть	čwert’	quarter	phon
74	vʃ	75	23,642	27.47	вшанувати	všanuvaty	to honor	morph
75	bʤ	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	zts	57	12,943	15.03	зцілення	zcilennja	healing	morph
82	tl	56	42,274	49.11	тлумачення	tlumačennja	translation	phon

	CLUSTER	LEMMA TYPES	TOKENS	FREQ PER MIL	UKRAINIAN	TRANSLITERATION	TRANSLATION	PHON /MORPH
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	глорія	glorija	glory	phon
88	gv	41	8,521	9.9	гвалт	gvalt	uproar	phon
89	ʧm	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	ʧl	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	tsm	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	львівский	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

	CLUSTER	LEMMA TYPES	TOKENS	FREQ PER MIL	UKRAINIAN	TRANSLITERATION	TRANSLATION	PHON /MORPH
104	tsn	11	5,638	6.55	цнота	cnota	virtue	phon
105	tx	9	1,712	1.99	тхір	txir	ferret	phon
106	ʤb	7	1000	1.16	дзбан	dzban	pitcher	phon
107	mʧ	7	16,552	19.19	мчати	mčaty	race	phon
108	ʤh	6	940	1.09	джгут	džxut	plait	phon
109	ʤh	6	2,060	2.39	джміль	džmil’	bumblebee	phon
110	rt	6	2,842	3.3	ртуть	rtut’	mercury	phon
111	ʧx	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	LEMMA TYPES	TOKENS	FREQ PER MIL	UKRAINIAN	TRANSLITERATION	TRANSLATION	PHON /MORPH
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	zgr	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

	CLUSTER	LEMMA TYPES	TOKENS	FREQ PER MIL	UKRAINIAN	TRANSLITERATION	TRANSLATION	PHON /MORPH
19	vsp	56	3,347	3.89	вспіти	v + spity	to be on time	morph
20	zfil	44	6,432	7.47	зглянутися	z + glianutysia	to take a look	morph
21	vfil	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	vfir	25	3,039	3.53	вгризатися	v + gryzatysia	to gnaw into	morph
29	vʃtʃ	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	zfin	17	1,679	1.95	згнити	z + gnyty	to rotten	morph

	CLUSTER	LEMMA TYPES	TOKENS	FREQ PER MIL	UKRAINIAN	TRANSLITERATION	TRANSLATION	PHON /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	vfin	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	zfv	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	vfv	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

	CLUSTER	LEMMA TYPES	TOKENS	FREQ PER MIL	UKRAINIAN	TRANSLITERATION	TRANSLATION	PHON /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	sʃtʃ	4	132	0.15	зщулився	s + ščulyty	to shrink	morph
57	vbl	4	445	0.52	вблагати	v + blagaty	to beg	morph
58	vtʃp	4	47	0.05	вшпарити	v + šparyty	to do sth energetically	morph
59	vtʃv	4	211	0.25	вчвал	včval	galloping	phon
60	stʃl	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	zdzv	2	56	0.07	здзвонитися	z + dzwonytysia	to call	morph
69	vtk	1	10	0.01	вткати	v + tkaty	to stick	morph