Linguistica Silesiana 40, 2019
ISSN 0208-4228

DOI: 10.24425/linsi.2019.129404

JOLANTA SZPYRA-KOZŁOWSKA
Maria Curie-Skłodowska University
jolanta.szpyra-kozlowska@poczta.umcs.lublin.pl

WHAT DO *KLEJ*, *LEK* AND *KULIG* HAVE IN COMMON? POLISH NATIVE SPEAKERS' JUDGEMENTS OF PHONOLOGICAL SIMILARITY BETWEEN WORDS AND NONWORDS

The paper approaches an important issue of the phonological similarity of words, relevant for current research in phonotactics, word recognition, production and acquisition, by analyzing the data collected in an experiment in which 30 native speakers of Polish were asked to provide phonologically similar words to 80 nonwords. The study demonstrates that the uncovered patterns of phonological similarity (segment substitutions, deletions and additions, the use of bigrams, trigrams and quadrigrams, noncontiguous sounds and segment metathesis) go beyond the commonly employed concept of *neighbourhood density* and point to the need to revise the current approaches to phonological similarity of words. It is argued that the experimental results can be attributed to the considerably more complex phonotactic and morphological structure of Polish than English.

Keywords: *phonological similarity between words*, *neighbourhood density*, *phonotactics*, *nonwords*

1. Introduction

One of the major concepts in current research on phonotactic knowledge is that of *neighbourhood density*, understood as the number of words which are phonologically similar to each other (e.g. Greenberg and Jenkins 1964, Vitevitch and Luce 1999, 2004, Bailey and Hahn 2001, Storkel 2013). It has been developed in the studies of English phonotactics and then shown to be relevant also for various other phenomena, such as word recognition, production and acquisition (Jusczyk 1997, Vitevitch et al. 1997, Storkel et al. 2010). While

an impressive body of research leaves no doubt that *neighbourhood density* is a significant aspect of linguistic knowledge, the notion has been applied mainly to English whose phonotactic structure is not very complex when compared to that of languages such as Polish which abound in complex consonant clusters found in all word positions. In other words, the provided evidence for the significance of the concept in question is mostly English-based and needs to be verified against relevant facts of other languages.

The present paper focuses on the issue of phonological similarity between words expressed in terms of *neighbourhood density* and its empirical verification. We report on an experiment in which 30 native speakers of Polish were given a list of 80 nonwords with well-formed phonotactic structure and requested to provide real Polish words phonetically and phonologically similar to the stimuli. This has been done with a view to finding answers to the following empirical and theoretical questions:

- What phonological similarity patterns between real words and nonwords do the experimental data reveal?
- Can they be accounted for in terms of *neighbourhood density*?
- What is the role of other factors (word length, segment similarity and orthography) in the participants' judgements?
- What theoretical implications does this research carry for the issue of phonological similarity?

To our knowledge, this is the first study devoted to the above issues based on Polish data.

The paper is structured as follows. Section 2 introduces some relevant theoretical concepts in phonotactic research and the issue of *neighbourhood density* in particular. Section 3 offers a brief overview of the major studies on Polish phonotactics. In Section 4 the experimental design is presented. Next the obtained results are provided and analysed. Section 5 deals with general results, Section 6 with the revealed similarity patterns between nonwords and real words and Section 7 addresses the role of other factors (word length, segment similarity and orthography) in the participants' judgements. The paper closes with a discussion and conclusions in Section 8.

2. Selected concepts and issues in phonotactic research

Phonotactics is usually defined (e.g. Trask 1996: 277) as "the set of restrictions in a language on the possible sequences of consonant and vowel phonemes within a word, a morpheme or a syllable."[1] In addition, phonotactics

[1] The units over which phonotactic constraints are defined depend on the adopted theoretical approach to phonology. For instance, in the traditional generative model of Chomsky and Halle (1968) they were expressed as *morpheme structure conditions*. In more recent frameworks (e.g. Hammond 2004) it is the syllable and its constituents over which phonotactic restrictions are

is concerned with the probability that a given segment or a sequence of segments will occur in a specific position within a syllable or word.

As is well-known, languages differ considerably in terms of their phonotactic structure, with those admitting only CV sequences placed at one end of the scale and those which contain heavy consonant clusters at the other end. While English and Polish, with numerous word-initial, medial and final consonant clusters, are located in the middle of this scale, the complexity of Polish phonotactics surpasses greatly that of English (for a comparison of the two systems see Zydorowicz *et al*. 2016). This means that many sequences of sounds which are possible (legal) in one language are impossible (illegal) in another. For instance, an English word may begin with three consonants which are sequences of /s/, followed by /p,t,k/ and then by /l,r,w,j/, e.g. *spring*, *street*, *sclerotic*, whereas in Polish various three and four-consonant clusters appear in this position (see the examples in the next section).

As often noted, phonemes found in a language can be combined in a large variety of ways, but only a small fraction of these combinations is employed in actually occurring words. Nonexistent sequences are divided into those which follow phonotactic constraints and are well-formed (termed accidental gaps or potential words) and those which violate phonotactic regularities (called systematic gaps). Both types of unattested items are referred to as nonwords, nonce words or pseudo-words[2] and play an important role in various psycholinguistic experiments (e.g. Vitevitch et al. 1997; Metsala 1997; Scarborough 2012; Storkel 2013), particularly those which involve decision tasks in which participants make judgements whether a given string of letters or sounds can become a real word of their native language (the issue of phonological 'goodness' of nonsense words) (on Polish native speakers' acceptability judgements of nonwords, see Szpyra-Kozłowska and Zydorowicz in press). Such ratings are claimed to be related to the phonological distance of a novel form from actual words, i.e. similarity between them.

The most frequently employed word similarity measure makes use of the concept of *phonological neighbourhood* (e.g. Greenberg and Jenkins 1964). It is defined (Frisch et al. 2004) as all words that are phonologically similar to a given sound sequence, i.e. those items which can be obtained from the source through a single sound substitution, deletion or addition. For instance, some neighbours of the word *cat* are *mat* and *bat* (substitution of the initial consonant) *cab* and *cap* (substitution of the final consonant) *caught* and *cot* (substitution of the vowel), *at* (initial consonant deletion) and *scat* (consonant addition). Those words which are similar to many other lexical items are said to have a *dense neighbourhood* and those which resemble few words have a *sparse*

---

defined. This issue plays no particular role in the present paper since we employ monosyllabic, monomorphemic nonwords only.

[2] Some researchers employ the term *pseudowords* to refer to legal nonwords, i.e. those which conform the phonological (and orthographic) patterns of a language.

neighbourhood. *Neighbourhood density* has been shown to play an important role not only in phonotactic judgements, but also in a variety of other phenomena. For example, dense neighbourhood facilitates word recognition and processing both in children and adults (Jusczyk 1997; Luce and Pisoni 1998; Grainger *et al*. 2005), affects the speed and accuracy of speech production as well as affects the frequency of speech errors (Vitevitch 2002).[3] Such phenomena are often referred to as neighbourhood-conditioned effects.

The term *wordlikeness* is often used to describe the similarity of nonwords to actual words. According to Bailey and Hahn (2001: 568), it is "the extent to which a sound sequence is typical of words in a language." Native speakers' wordlikeness judgements are argued to depend on the similarity between a nonword and words in the lexicon (*neighbourhood density*) as well as on the phonotactic structure of the nonword (phonotactic grammar). Both factors constitute elements of native speakers' phonological awareness and interact in phonotactic ratings (Vitz and Winkler 1973).

Moreover, as shown by Frisch *et al*. (2004) wordlikeness judgements are significantly influenced by the frequency of occurrence of sounds and sound sequences in the lexicon, i.e. their *phonotactic probability*. What is of much importance is also *bigram* frequency and *trigram* frequency.[4] The former concerns sequences of two contiguous elements while the latter three elements. *Phonotactic probability* and *neighbourhood density* are positively correlated as common segments and sequences of segments tend to occur in many similar sounding words. They were demonstrated to affect the recognition, production and acquisition of both real words and nonwords in a large variety of tasks, across different age groups and types of stimuli, both in native language and in foreign languages (Vitevitch and Luce 1999; Storkel 2013).

The developments sketched above led to the rise of several sophisticated probabilistic models of phonotactics (e.g. Coleman and Pierrehumbert 1997; Bailey and Hahn 2001; Vitevitch and Luce 2004; Frisch et al. 2004, a maximum entropy model of Hayes and Wilson (2008)), which, for reasons of length limitations and little relevance to our study with its focus on phonological similarity of words, will not be discussed here.

[3] According to Vitevitch (2002), in tongue twisters more errors are made in words with sparse neighbourhood than in items with dense neighbourhood.

[4] Unigrams, bigrams and trigrams play a crucial role in the so-called *n*-Gram models of phonotactics, widely used in psycholinguistics and natural language processing (Kager and Pater 2012). They are based on the frequency of segments and sequences of segments of length *n* on the words of a language. The rarity of strings with *n* bigger than two means that usually only unigrams and bigrams are employed and longer sequences are treated as combinations of shorter ones.

3. Basic facts on Polish consonant clusters and phonotactic research

In this paper the focus is on the issue of similarity between Polish real words and nonwords involving two-consonant clusters. It is therefore necessary to present some basic facts on consonant sequencing in Polish and phonotactic research devoted to it.

While there are some restrictions on the sequences of consonants and vowels, of only consonants and only vowels,[5] the most striking feature of Polish phonotactics is the fact that it allows for many different consonant clusters in all positions. Thus, up to four consonants[6] can be found word initially and up to five consonants word finally, as shown in (1).

(1)

examples of words with initial clusters:	examples of words with final clusters:
drgnąć [drgnɔɲtɕ] ‘budge’	*barszcz* [barʂtʂ]’borscht’
pstry [pstrɨ] ‘multicoloured’	*blichtr* [blʲixtr] ‘sham’
źdźbło [ʑdʑbwɔ] ‘stalk’	*głupstw* [gwupstf]
wstrzelić [fstʂɛlʲitɕ] ‘shoot’	*następstw* [nastɛmpstf] ‘results, gen.’

Such clusters represent a large variety of segmental combinations and appear to follow no surface-true rules. Zydorowicz *et al*. (2016: 74) in the corpus of 48.6 million words,[7] found 2451 cluster types including 454 initial, 1793 medial and 204 different consonant sequences. In the corpus of 410 million words of American English (Corpus of Contemporary American English) they found 861 cluster types including 55 initial, 687 medial and 119 final consonant sequences. These figures show a considerable difference in the consonant clustering possibilities of the two languages.

The complexity of consonant sequences in Polish has raised much interest among phonologists, as seen in numerous studies devoted to this issue which appeared both in the previous century (e.g. Bargiełówna (1950); Kuryłowicz (1952); Ułaszyn (1956); Kreja (1969); Leszczyński (1969); Rocławski (1981); Laskowski (1985); Dunaj (1985); Dobrogoska (1992) as well as in the present one, e.g. by Śledziński (2010); Jaskuła and Szpyra-Kozłowska (in press)), approaching it from a variety of perspectives and covering various aspects of clusters: diachronic changes, their synchronic status and structure both in Standard Polish and in local dialects.

---

[5] For instance, palatal and palatalized consonants are followed by the high front vowel [i], the remaining ones by the high centralized [ɨ], clusters of obstruents must be uniform in voicing and no two vowel sequences are found in single native morphemes.

[6] The list of Polish consonants is provided in Appendix 1.

[7] The corpus was the full text of the *Rzeczpospolita* newspaper from the 2000-2001 period.

The studies of Polish consonant clusters adopt different theoretical stances. Many of those mentioned above are analyses carried out mainly within structural, Prague-type linguistics. Other publications were offered within generative models, in reference to syllable structure (e.g. Bethin 1992, Rubach and Booij 1990, Szpyra 1995), in Government Phonology (e.g. Cyran and Gussmann 1998), Optimality Theory (Rochoń 2000) and Natural Phonology (e.g. Dziubalska-Kołaczyk 2014; Zydorowicz *et al*. 2016).

As these approaches differ considerably in terms of their theoretical assumptions, degree of analytic abstractness and the employed descriptive mechanisms, their presentation and comparison is beyond the scope of this paper. Suffice it to say that no agreement has been reached as to how Polish consonant clusters and their structure should be adequately described and interpreted.

As this brief introduction to Polish phonotactics suggests, this language, with its complex consonantal structure, departs considerably from phonologically simpler languages such as English and provides an excellent testing ground for various theoretical concepts such as *neighbourhood density*.

4. Experimental design

Below we characterize the experimental design in terms of its participants, the stimuli and the adopted procedure.

4.1. The participants

The participants were a group of 30 4th year students (all females) of the speech therapy department of Maria Curie-Skłodowska University, Lublin, aged 22-23, with a fairly good knowledge of Polish phonetics, but no earlier training in phonotactics.

4.2. The stimuli

The stimuli were 80 monosyllabic nonwords,[8] all with phoneme sequences attested in real Polish words, i.e. constituting the so-called accidental gaps. The stimuli were taken from a different study (Szpyra-Kozłowska and Zydorowicz in press), meant to examine Polish native speakers' acceptability judgements of nonwords with two-consonant clusters. Thus, 33 items contained word initial consonant sequences and 35 word-final clusters of different frequency. (12 forms were fillers of the following structure: 6 CVC, 2 CCCVC, 1 CCVCC,

[8] The list of stimuli is given in Appendix 2.

3 CVCCC.[9] All the items were included in the present analysis. The use of the same stimuli was motivated by the wish to compare the results of both studies.[10] The choice of monosyllabic items was also dictated by the need to avoid additional issues involved in word length and division into syllables.

4.3. The procedure

The participants were asked to complete an anonymous questionnaire in which they were requested to provide real Polish words phonetically similar to the 80 nonwords. Although the stimuli appeared in the written form, the students were supposed to pronounce each item before they made their decisions.[11] They were encouraged to supply as many similar words as they could think of and were requested not to consult the others. It was explained to them that there were no wrong answers as they were all equally competent native speakers of Polish. There was no time limit to complete the questionnaires. It took the subjects between 30 and 45 minutes to carry out the task. The goals of the experiment were not explained to them.

It is worth adding that the students participated in the experiment willingly and displayed much interest in it as shown in a lively discussion which followed its completion.

5. General results

The participants provided 1850 real words judged by them similar to the experimental stimuli, many of which were the same. About 25 tokens had to be rejected either because they could not be deciphered or because they were different inflectional forms of one word. This means that the analysis below includes 1825 items.

Some respondents provided more real words than one (from 2 to 4) for a nonword. Frequently, however, no answers were given at all. The average number of the supplied responses by all the participants was 35 per a stimulus, with the lowest being 29 and the highest 119. Thus, in completing the task the nonwords presented to the subjects a different degree of difficulty. The easiest items in terms of providing similar real words to them, turned out to be *luść* [luɕtɕ], *człać* [tʂwatɕ], *wikr* [vʲikr], *siap* [ɕap], *plaj* [plaj], *gedź* [gɛdʑ], *żaft* [ʐaft],

[9] It is worth adding that the majority of studies devoted to neighbourhood effects in English employs items of CVC structure which contains no consonant clusters.

[10] This is done in a forthcoming publication.

[11] In written tasks we can talk about orthotactics, i.e. written phonotactics. Orthographic neighbours are defined as the number of words that can be formed from a given item by changing a letter. The impact of orthography on the experimental results is discussed in some detail in Section 7.

ciusk [tɕusk], *szruń* [ʂruɲ], *kleg* [klɛk], *stryw* [strɨf], *preń* [prɛɲ], *lik* [lʲik], *styś* [stɨɕ], *drecz* [drɛtʂ], *ftać* [ftatɕ] and *trzeg* [tʂɛk], with the "winner" being *drecz* (64 answers). The most difficult nonwords, which yielded only 10-15 tokens, with frequently no response from many students are as follows: *żlorz* [ʐlɔʂ], *dnysz* [dnɨʂ], *słemf* [swɛmf], *somf* [sɔmf], *binf* [bʲnf], *czwyp* [tʂfɨp], *sełsz* [sɛwʂ], *julm* [julm], *fupr* [fupr], *wniup* [vɲup] and *rojp* [rojp], with *fupr* obtaining only 10 answers.

An analysis of the above examples shows that no strong correlation is found between *neighbourhood density* of nonwords and the number of the provided responses. For instance, only a few respondents (4) supplied the word *dysz* [dɨʂ] 'nozzles, gen.,' which is minimally different from the nonword *dnysz* [dnɨʂ] (the second group) while *szruń* (the first group), with no minimal pair correspondents, inspired the majority of the participants (24) to provide such items as *szron* [ʂron] 'frost,' *sznur* [ʂnur] 'rope,' *szary* [ʂarɨ] 'grey' and *szuruj* [ʂuruj] 'shuffle.' This means that in their decisions concerning phonologically similar and dissimilar words the subjects must have been guided by other criteria than *neighbourhood density*. They are discussed in the next section.

6. Patterns of phonological similarity between nonwords and real words

In this section we examine the major patterns of phonological similarity found between the stimuli and the lexical items provided by the participants.

Single segment substitution

Recall that *neighbourhood density* involves items which differ in terms of single segments. Within the experimental data we find numerous cases of substitutions creating minimal pairs. They involve initial (1a), final (1b) and medial (1c) consonants as well as vowels (1d):[12]

(2)

a. substitution of the initial consonant:

luść [luɕtɕ] – ***p***uść [puɕtɕ] 'let go'	***cz***łać [tʂwatɕ] – ***p***łać [pwatɕ] 'pay'
gedź [gɛtɕ] – ***si***edź [ɕɛtɕ] 'sit'	***g***edź [gɛtɕ] – ***j***edź [jɛtɕ] 'go'
żaft [ʐąft] – ***h***aft [xaft] 'embroidery'	***dż***acht [dʐąxt] – ***j***acht [jaxt] 'yacht'
cior [tɕɔr] – ***p***or [pɔr] 'leak'	***t***rzeg [tʂɛk] – ***b***rzeg [bʐɛk] 'coast'

[12] The transcription employed in this paper reflects some common phonological processes of Polish, such as word final obstruent devoicing, palatalization of consonants before the following /i/ and /j/, and voice assimilation of adjacent obstruents. These modifications are responsible for some spelling-pronunciation discrepancies. It should be added, however, that in all instances the written form of the stimuli corresponds unambiguously to the same pronunciation.

b. substitution of the final consonant:

*cio**r*** [tɕɔr] – *cio**s*** [tɕɔs] ‘blow’	*kle**g*** [klɛk] – *kle**j*** [klɛj] ‘glue’
*wik**r*** [vʲikr] – *wik**t*** [vʲkt] ‘food’	*trze**g*** [tʂɛk] – *trze**ć*** [tʂɛtɕ] ‘rub’
*gła**l*** [gwal] – *gła**z*** [gwas] ‘stone’	*spy**p*** [spɨp] – *spy**ch*** [spɨx] ‘push’
*pre**ń*** [prɛŋ] – *pre**cz*** [prɛtʂ] ‘go away’	*gu**f*** [guf] – *gu**z*** [gus] ‘bump’

c. substitution of a medial consonant:

*pł**a**j* [plaj] – *pch**a**j* [pxaj] ‘push’	*ża**f**t* [ʐaft] – *ża**r**t* [ʐart] ‘joke’

d. vowel substitution:

*l**u**ść* [luɛtɕ] – *l**i**ść* [lʲiɛtɕ] ‘leaf’	*l**i**k* [lʲik] – *l**e**k* [lɛk] ‘drug’
*w**u**lc* [vults] – *w**a**lc* [valts] ‘waltz’	*f**y**ks* [fɨks] – *f**a**ks* [faks] ‘fax’
*g**u**f* [guf] – *g**a**f* [gaf] ‘blunders, gen.’	*str**y**w* [strɨf] – *str**a**w* [straf] ‘foods, gen.’
*dr**e**cz* [drɛtʂ] – *dr**o**cz* [drɔtʂ] ‘bicker’	*ci**a**k* [tɕak] – *ci**e**k* [tɕɛk] ‘flow’

It should be noted that while substitutions of word initial and final consonants as well as vowels are often found in the data, cases of the modification of medial (nonperipheral) consonants in (2c) are infrequent. This is in agreement with an observation of the perceptual salience of word initial and final consonants, but not word internal consonants (Copeland and Radvansky 2001).

Single segment deletion

In many instances the provided real words differ from nonwords in terms of a segment deleted from the stimulus. This can be either one of consonants in the initial cluster (3a) or the final cluster (3b).

(3)

a. consonant deletion in the initial cluster:

*d**n**ysz* [dnɨʂ] – *dysz* [dɨʂ] ‘nozzles, gen.’	*s**p**yp* [spɨp] – *syp* [sɨp] ‘build’
*g**ł**al* [gwal] – *gal* [gal] ‘galas, gen.’	*ch**ł**uf* [xluf] – *luf* [luf] ‘barrels, gen.’
*k**ł**eg* [klɛk] – *keg* [kɛk] ‘keg’	*p**r**eń* [prɛŋ] – *pień* [pʲjɛŋ] ‘trunk’

b. consonant deletion In the final cluster:

*ce**k**l* [tsɛkl] – *cel* [tsɛl] ‘goal’	*dy**l**sz* [dɨlʂ] – *dysz* [dɨʂ] ‘nozzles, gen.’
*ru**dź**m* [rudʑm] – *rum* [rum] ‘rum’	*for**s**t* [fɔrst] – *fort* [fɔrt] ‘fort’

In the above examples the consonants which are dropped are found both in peripheral and nonperipheral positions.[13]

Single segment addition

Less frequent cases involve segment addition with either vowels (4a) or consonants (4b) being added to nonwords to create real words, e.g.

[13] No vowel deletion takes place in the experimental items as these are all monosyllabic forms which require a nucleus and only vowels can have this function in Polish.

(4)

a. vowel addition:

lżag [lzągk] – *le**ż**ak* [lɛzągk] ‘deckchair’	*łzur* [wzur] – *l**a**zur* [lazur] ‘azure’
rwol [rvɔl] – *r**y**wal* [rɨval] ‘rival’	*ksztaf* [kştaf] – *k**a**sztan* [kaştan] ‘chestnut’

b. consonant addition:

drecz [drɛtş] – *dre**s**zcz* [drɛştş] ‘shiver’	*ftać* [ftatɕ] – *w**s**tać* [fstatɕ] ‘get up’

The transformations illustrated in (4a) and (4b) are not numerous in the experimental data and are less frequent than modifications presented earlier. Moreover, in order to provide real Polish words, the participants sometimes introduced some additional changes and departed from the original nonwords. For example in the pair *łzur* [wzur] – *lazur* [lazur] the initial glide is replaced with the lateral, in *rwol* [rvɔl] – *rywal* [rɨval] /o/ is replaced with /a/, in *ksztaf* [kştaf] – *kasztan* [kaştan] the final /f/ is substituted with /n/. In other words, these are not exact minimal pairs, but near-minimal pairs.

To sum up the results presented so far, instances of segment substitutions which give rise to minimal pairs appear to outnumber cases of segment deletion, while segment addition is the least frequently employed modification.

Let us now present some examples which involve more radical departures in the structure of real words the participants considered similar to the experimental nonwords. Put differently, we will now examine those cases which cannot be subsumed under the traditional definition of *neighbourhood density*.

Bigrams

A large class of the provided items is characterized by the presence of the same bigrams in the source items and the proposed actual words. They concern two initial (5), two final (7) and two middle (6) segments.

(5) initial bigrams

a. #CC:

***lż**ag* [lzągk] – ***lż**ej* [lzęj] ‘lighter’	***ml**on* [mlɔn] – ***ml**eć* [mlɛtɕ] ‘grind’
***łz**ur* [wzur] – ***łz**y* [wzɨ] ‘tears’	***chl**uf* [xluf] – ***chl**eb* [xlɛp] ‘bread’

b. #CV:

***ni**dm* [ɲidm] – ***ni**kt* [ɲikt] ‘nobody’	***cio**pń* [tɕɔpɲ] – ***cio**s* [tɕɔs] ‘blow’
***li**k* [lʲik] – ***li**tr* [lʲitr] ‘litre’	***lu**ść* [luɕtɕ] – ***lu**fa* [lufa] ‘barrel’

In (5a) the bigrams comprise two consonants and in (5b) the initial CV sequences.

(6) final bigrams
a. CC#:
*żem**s*** [zɛ̨ms] – *gzy**ms*** [gzɨms] ‘cornice’ *tu**ś**ń*** [tuɛ̯n] – *pie**ś**ń*** [pʲjɛɛ̯n] ‘song’
*ka**jcz*** [kajtʂ] – *pe**jcz*** [pɛjtʂ] ‘whip’ *bo**lp*** [bɔlp] – *pu**lp*** [pulp] ‘pulps, gen.’
b. VC#:
*mn**ep*** [mnɛp] – ***lep*** [lɛp] ‘flypaper’ *rdz**up*** [rdzup] – ***zup*** [zup] ‘soups, gen.’
*dn**ysz*** [dnɨʂ] – ***mysz*** [mɨʂ] ‘mouse’ *g**edź*** [gɛdʑ] – *ś**ledź*** [ɕlɛtɕ] ‘herring’

The examples in (6a) share the final CC clusters while those in (6b) the final VC sequences.

(7) middle bigrams (CV, VC)
*pr**zu**n*** [pʂun] – *br**zuc**h*** [bzux] ‘belly’ *s**ejsz*** [sɛjʂ] – *r**ejs*** [rɛjs] ‘cruise’

It is striking that while initial (5) and final (6) bigrams are numerous, middle bigrams (7) are infrequent, in agreement with the claim of a greater salience of word initials and finals than medials.

Trigrams

Let us now deal with trigrams found in the items considered similar.[14]

(8) initial trigrams
a. #CVC:
szuk**t [ʂukt] – ***szuk**a*** [ʂuka] ‘he searches’ ***bin**f*** [bʲinf] – ***bin**go*** [bʲiŋgɔ] ‘bingo’
wik**r [vʲikr] – ***Wik**tor*** [vʲktɔr] ‘Victor’ ***kaj**cz*** [kajtʂ] – ***kaj**ak*** [kajak] ‘canoe’
b. #CCV:
trz**eg [tʂɛk] – ***trz**eba*** [tʂɛba] ‘should’ ***czł**ać*** [tʂwatɕ] – ***czł**apać*** [tʂwapatɕ] ‘plod’
skor**z [skɔʂ] – ***skor**o*** [skɔrɔ] ‘since’ ***chluf*** [xluf] – ***chlu**pać*** [xlupatɕ] ‘splash’

Within this type in each pair we can find the same CVC sequences (8a) and CCV strings (8b).

Several final trigrams involving CVC and VCC sequences also occurred in our data, the majority containing VCC sequences. No middle trigrams were found.

[14] It should be added that the majority of forms in (2)-(4), i.e. minimal pairs also contain bigrams and trigrams.

(9) Final trigrams

a. CVC#:

drecz [drɛtʂ] – *strecz* [strɛtʂ] ‘stretch’ ***cior*** [tɕɔr] – *bu**cior*** [butɕɔr] ‘shoe, augm.’

b. VCC#:

żaft [ʐaft] – *kr**aft*** [kraft] ‘craft’ ***łzur*** [wzur] – *gla**zur*** [glazur] ‘glaze’

*ci**usk*** [tɕusk] – *pl**usk*** [plusk] ‘splash’ ***kefl*** [kɛfl] – *tr**efl*** [trɛfl] ‘clubs’

Quadrigrams

In a few instances we noted the presence of quadrigrams whose number is very limited due to the monosyllabic character of the experimental nonwords, e.g.

(10)

a. initial quadrigrams:

kształ [kʂtaw] – ***kształt*** [kʂtawt] ‘shape’ ***plaj*** [plaj] – ***plajta*** [plajta] ‘bankruptcy’

kostw [kɔstf] – ***kostka*** [kɔstka] ‘cube’ ***luść*** [luɕtɕ] – *czeluść* [tʂɛluɕtɕ] ‘abyss’

b. final quadrigrams:

skorz [skɔʂ] – ***skorzystać*** [skɔʐɨstatɕ] ‘use’ ***skorz*** [skɔʂ] – *pi**skorz*** [pʲskɔʂ] ‘weatherfish’

In (10a) the initial quadrigrams are employed whereas in (10b) final quadrigrams, with the former being more numerous than the latter.

Two noncontiguous segments

The next group of items involves the use of two noncontiguous segments.

(11)

dżyml [dʐɨml] – ***dżem*** [dʐɛm] ‘jam’ ***rudźm*** [rudʑm] – ***rodzić*** [rɔdʑitɕ] ‘give birth’

szojf [ʂɔjf] – ***szef*** [ʂɛf] ‘boss’ ***tkuf*** [tkuf] – ***tłuc*** [twuts] ‘break’

In (11) two noncontiguous sounds, usually consonants, of the nonwords are used in real words.

Three and four segments

Numerous cases involve the employment of bigrams and another sound separated from the former with some other segment (12a), three segments forming bigrams and trigrams in the original items and used contiguously or not in the real words (12b) and a similar type of pattern but including four sounds in (12c).

(12)

a.

szobl [ʂɔbl] – ***szabla*** [ʂabla] ‘sword’ ***kajcz*** [kajtʂ] – ***kaczka*** [katʂka] ‘duck’
rwol [rvɔl] – ***rywal*** [rɨval] ‘rival’ ***człać*** [tʂwatɕ] – ***czekać*** [tʂɛkatɕ] ‘wait’

b.

wulc [vults] – ***widelec*** [vʲidɛlɛts] ‘fork’ ***ciak*** [tɕak] – ***cielak*** [tɕɛlak] ‘calf’
gechć [gɛxtɕ] – ***grzech*** [gzʐex] ‘sin’ ***wikr*** [vʲkr] – ***iskra*** [iskra] ‘spark’

c.

forst [fɔrst] – ***porost*** [pɔrɔst] ‘growth’ ***kastrz*** [kastʂ] – ***korsarz*** [kɔrsaʂ] ‘privateer’

What all the items in (12) have in common, both actual words and nonwords, is the fact of sharing by them three or four sounds in different configurations presented above. All the relevant segments, however, appear in the same order in the stimuli and the real lexical items.

Segment metathesis

The most interesting cases, however, are those in which a real word provided by the participants as similar to a given nonword involves the reordering (metathesis) of some of the segments found in the original items, i.e. two in (10a), three in (10b) and four in (10c).

(13)

a.

lik [lʲik] – ***kit*** [cit] ‘putty’

b.

guf [guf] – ***fuga*** [fuga] ‘fugue’ ***nidm*** [ɲidm] – ***dni*** [dɲi] ‘days’
gedź [gɛtɕ] – ***gdzie*** [gdʑɛ] ‘where’ ***fupr*** [fupr] – ***puf*** [puf] ‘pouffe’
mnep [mnɛp] – ***menel*** [mɛnɛl] ‘tramp’ ***siap*** [ɕap] – ***pasi*** [paɕi] ‘OK’
bolp [bɔlp] – ***plomba*** [plɔmba] ‘filling’ ***słemf*** [swemf] – ***hełm*** [xɛwm] ‘helmet’
szorl [ʂɔrl] – ***szron*** [ʂrɔn] ‘frost’ ***plaj*** [plaj] – ***Alp*** ‘Alps, gen. pl.’
siap [ɕap] – ***psia*** [pɕa] ‘dog, adj.’ ***kefl*** [kɛfl] – ***klif*** [klʲif] ‘cliff’

c.

ciopń [tɕɔpɲ] – ***nicpoń*** [ɲitspɔɲ] ‘wastrel’ ***gełc*** [gɛwts] – ***cegła*** [tsɛgwa] ‘brick’
przun [pʂun] – ***żupan*** [ʐupan] ‘dress’[15] ***szukt*** [ʂukt] – ***sztuk*** [ʂtuk] ‘pieces, gen.’
kefl [kɛfl] – ***flek*** [flɛk] ‘heel tip’ ***szruń*** [ʂruɲ] – ***sznur*** [ʂnur] ‘rope’

[15] Traditional dress of Polish noblemen.

Among instances of segment reordering we find many items involving three (12b) and four (12c) sounds, and only one with two such segments (12a).

It is now important to establish the frequency with which the presented similarity patterns were employed by the participants. Towards this purpose, below we examine the segmental structure of the most frequently provided words for a given stimuli. In 9 cases no dominant item was found,[16] in the remaining instances one, two or three such words can be indicated. 103 most frequently provided lexical items were found in our data. Below they are divided into several categories depending on the employed similarity pattern.

(13)
a. substitutions, deletions and additions – 27
b. bigrams – 24
c. trigrams and quadrigrams – 17
d. bigrams + a noncontiguous segment – 15
e. 2-3 noncontiguous segments – 14
f. segment reordering – 5
g. trigrams + a noncontiguous segment – 1

Among the most frequent words provided by the participants the two largest groups involve single segment substitutions (with one case of deletion and one of insertion) and bigrams (13a, 13b). The next big classes include trigrams and quadrigrams (13c), bigrams with a noncontiguous segment (13d) and words with 2 or 3 noncontiguous segments (13e). The remaining patterns are less frequent. It is worth adding that within 103 items in (13) 60% share 3 segments with the original nonwords and 40% 2 segments. The same observation holds true in the case of all the forms which have at least two or three sounds in common, regardless of whether these segments are contiguous or not, and regardless of their order. What matters here is the presence of some identical sounds which is sufficient for considering two items as similar.[17]

7. Other factors in similarity judgements: word length, segment resemblance and orthography

It is also important to examine briefly the role of several other potential determiners of the participants' similarity judgements. The first issue at stake is whether the items viewed as similar are of the same length, measured in terms of the number of syllables they contain. Recall that the stimuli were all monosyllables. As the examples below demonstrate, indeed many real Polish

[16] These are the following nonwords: *człać*, *kastrz*, *binf*, *czwyp*, *mloń*, *gełc*, *sełsz*, *kajcz*, *szorl*.

[17] The number of shared sounds is probably connected with the length of the original forms and the number of phonemes they contain. In order to draw more definitive conclusions in this respect, a study with other stimuli is needed.

words supplied by the students were also monosyllabic, but many of them were longer. In (14) we present a selection of the provided items which consists of single syllables (14a), two syllables (14b) and three syllables (14c).[18]

(14)

a.	*żaft* [ʐaft] – *żart* [ʐart] ‘joke’	*wikr* [vʲikr] – *wir* [vʲir] ‘whirl’
	preń [prɛɲ] – *cień* [tɕɛɲ] ‘shadow’	*wypt* [vɨpt] – *szept* [ʂɛpt] ‘whisper’
b.	*dzylsz* [dzɨlʂ] – *dyszeć* [dɨʂɛtɕ] ‘pant’	*szruń* [ʂruɲ] – *szuruj* [ʂuruj] ‘shuffle’
	spyp [spɨp] – *sypać* [sɨpatɕ] ‘pour’	*fyks* [fɨks] – *feniks* [fɛɲiks] ‘phoenix’
c.	*cior* [tɕɔr] – *cieciorka* [tɕɛtɕɔrka] ‘chick peas’	*dzbyw* [dzbɨf] – *zdobywca* [zdɔbɨftsa] ‘winner’
	ftać [ftatɕ] – *haftować* [xaftovatɕ] ‘embroider’	*skorz* [skoʂ] – *skorzystać* [skoʐɨstatɕ] ‘use’

We can conclude that word length does not play a crucial role in similarity judgements and what matters is mainly the presence of the same sounds and sound sequences in the involved items. To be more precise, when monosyllabic real words resembling the original nonwords were available, the participants often made use of them. This, however, was not always possible due to the fact that Polish, as an inflectional language which usually requires the presence of inflectional affixes, does not contain as many monosyllabic words as English. For instance, if a given nonword began in a similar way to some verb, a verbalizing suffix (e.g. *-ać*, *-eć*, *-ować*) had to be employed making thus the original form longer, e.g., *wypt* [vɨpt] – *wypytać* [vɨpɨtatɕ] ‘ask.’

Let us now address briefly the issue of sound similarity in pairs of nonwords and real words. In many cases it can be argued that nonidentical segments appeared in the experimental data due to their perceived resemblance to some other sounds. Below we present some relevant examples which can be attributed to consonant similarity.

(15)

a.	palatals and nonpalatals	
	/n – ɲ/: *dnysz* [dnɨʂ] – *dni* [dɲi] ‘days’	*mnep* [mnɛp] – *mnie* [mɲɛ] ‘me’
	/m – mʲ/: *żmyg* [ʐmɨk] – *żmij* [ʐmʲij] ‘vipers, gen.’	*ćmesz* [tɕmɛʂ] – *śmiesz* [ɕmʲjeʂ] ‘you dare’
	/p – pʲ/: *preń* [prɛɲ] – *pień* [pʲɛɲ] ‘trunk’	*ciopń* [tɕɔpɲ] – *pień* [pʲɛɲ] ‘trunk’
	/l – lʲ/: *julm* [julm] – *Julia* [julʲja] ‘Julia’	*bolp* [bɔlp] – *boli* [bɔlʲi] ‘it hurts’
	/s – ɕ/: *żems* [ʐɛms] – *żeś* [ʐɛɕ] ‘that you’	*siamn* [ɕamn] – *sam* [sam] ‘alone’
b.	voiced and voiceless obstruents	
	/pʲ – bʲ/: *binf* [bʲinf] – *pin* [pʲin] ‘PIN’	/ʂ – ʐ/: *lżag* [lʐak] – *szlag* [ʂlak] ‘be over’
	/f – v/: *ftać* [ftatɕ] – *witać* [vʲtatɕ] ‘greet’	/ɕ – ʑ/: *tuśń* [tuɕɲ] – *tuzin* [tuʑin] ‘dozen’

[18] No words longer than three syllables were found in the data.

In (15) the stimuli share with the provided words the presence of the consonants which differ with regard to palatality (nonpalatal versus palatal / palatalized segments) in (15a) while those in (15b) with respect to obstruent voicing (voiced and voiceless). The choice of the real words could be attributed to the phonetic similarity of the consonants in question.

As mentioned earlier, while the participants were asked to take into account the pronunciation of the stimuli, the written form could have affected their decisions as well. An examination of the experimental data demonstrates that in many cases orthography was not taken into account since segments spelt in two different ways but pronounced identically were often found in pairs of items regarded as similar. Selected examples are provided in (16).

(16)

a.	*prz**u**n* [pʂun] – *prz**ó**d* [pʂut] ‘front’	*ł**z**u**r*** [wzur] – *w**z**ó**r*** [vzur] ‘pattern’
b.	***ż**myg* [ʐmɨk] – ***rz**emyk* [ʐɛmɨk] ‘strap’	***ż**ems* [ʐɛms] – ***rz**ęs* [ʐɛ̃ws] ‘eyelashes, gen.’
c.	*żlo**rz*** [ʐlɔʂ] – *klo**sz*** [klɔʂ] ‘lampshade’	*ge**dź*** [gɛtɕ] – *le**ć*** [lɛtɕ] ‘fly’

In (16) we can find pairs of words spelt in two ways: with [u] written as <u> and <ó> in (16a), [ʐ] spelt as <ż> and <rz> in (16b) and word final (phonetically voiceless) obstruents written as voiced or voiceless in (16c). Such items suggest that the participants’ decisions were based on phonetic rather than orthographic shapes of words.

In several examples, however, the impact of spelling could be noted.

(17)

a. ***dzb**yw* [dzbɨf] – ***zb**yt* [zbɨt] ‘too,’ ***zb**aw* [zbaw] ‘save’
b. ***dzy**lsz* [dziłʂ] – ***dy**sza* [dɨʂa] ‘nozzle,’ ***dy**cha* [dɨxa] ‘ten’

In (17) the initial voiced dental affricate [dz] in the two nonwords (spelt with two letters <dz>), in real words appears either as the corresponding fricative [z] in (17a) or the plosive [d] in (17b), which can be attributed to spelling-based similarity between all three consonants. It should be added that these are isolated examples in our data, which means that in the majority of cases in their decisions the participants were guided mainly by word pronunciation.

We can conclude this section by claiming that the participants’ judgements were not significantly affected by the length and spelling of words, but were influenced by the phonetic similarity between sounds. More evidence, however, is needed to settle these issues satisfactorily.

8. Discussion and conclusions

In the majority of studies devoted to the question of phonological word similarity, this notion is usually defined in terms of *neighbourhood density* which involves counting the number of minimal pairs that can be formed from the source item. While this approach is workable with languages with relatively simple phonotactic grammar and numerous short words, such as English, it is, as shown in this paper, of limited applicability in the case of Polish with its large number and variety of consonant clusters, and the prevalence of polysyllabic words.

Thus, the participants viewed two items as similar not only when a nonword and a real word formed a minimal pair (through segment substitution, deletion or addition), but also in several other cases involving bigrams, trigrams and quadrigrams, the presence of noncontiguous segments, noncontiguous segments and bigrams, and by sound reordering. The employment of several of these patterns is reflected in the title of the paper with the words *klej* [klɛj] 'glue,' *lek* [lɛk] 'medicine' and *kulig* [kulʲik] 'sleigh ride,' which were all provided as similar to the nonword *kleg* [klɛk], with *klej* and *lek* forming minimal pairs with the source item (through the substitution of the final consonant and through the deletion of the initial segment respectively), and *kulig* representing the use of three original consonants with vowels inserted between them.

It might be hypothesized that Polish native speakers, when faced with the task of suggesting real words similar to nonwords, reach for the available minimal pairs, but in their absence (or when none of them is recalled at a given moment), make use of other patterns listed above. This is often shown in several different proposals made by a single participant. One of them, for instance, offered the following similar words to *plaj* [plaj]: *pluj* [pluj] 'spit' (vowel substitution), *pled* (the initial bigram) and *polej* (three noncontiguous consonants). Another subject's proposals of words similar to the same item are *plac* (substitution of the final consonant), *placz* [pwatʃ] 'cry' (the use of two noncontiguous segments) and *plajta* [plajta] 'bankruptcy' (initial quadrigram and an additional syllable). Yet a different student's suggestions for this nonword are *pchaj* [pxaj] 'push' (consonant substitution), *maj* [maj] 'May' (the final bigram) and *plejada* [plɛjada] 'array' (the initial bigram, a noncontiguous consonant and two added syllables). Two conclusions can be drawn from such data. First, the fact that individual participants frequently provided more than one actual word for a given stimulus shows that several patterns of similarity are employed by one native speaker. They also indicate that in making similarity judgements the participants took the whole words into account and not only selected segments or their sequences in particular word positions.

What the experimental material collected in this paper has demonstrated is that subjective similarity judgements between nonwords and actual Polish words concerned forms with at least two identical segments, both contiguous and noncontiguous. In many cases three or four identical sounds were present

in pairs of items viewed as similar. Of the three possible determinants of the participants' decisions, word length and orthography turned out to be of minor importance while phonetic similarity between phonemes was shown to play an important role in their proposals.

In the previous pages *neighbourhood density* has been argued to be insufficient for Polish as a reliable measure of native speakers' subjective word similarity judgements. In this respect we provide compelling evidence for the correctness of Bailey and Hahn's (2000) criticism of this concept, who point out that it fails to take into account phonetic similarity between phonemes. For example, replacing /b/ with /p/, which differ only in terms of voicing (e.g. in *bat – pat*), yields a neighbour, just like replacing /b/ with /s/, differing with regard to the place and manner of articulation (e.g. in *bat – sat*). Moreover, the approach relying only on the notion in question ignores all words which do not form minimal pairs with the base items. As shown above, Polish provides ample evidence that this criticism is fully justified. Thus, we agree with Bailey and Hahn (p. 572) that "more sophisticated measures of word similarity are required if we wish to acquire more than a superficial understanding of neighbourhood effects."[19] Such measures should be able to cover all the similarity patterns uncovered in this paper.

Since this is the first study on Polish native-speakers' judgements of phonological similarity between nonwords and real words, with a limited number and types of stimuli and only 30 participants, the presented conclusions should be verified against a larger body of empirical data. Nevertheless, in spite of all these limitations, the present paper provides ample evidence that the current approaches to word similarity relying heavily on the concept of *neighourhood density* are too simplistic and in need of substantial revision when faced with languages, such as Polish, with more complex phonotactic structure than English.

References

Bailey, T.M., and U. Hahn 2001. Determinants of wordlikeness: phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44: 568-591.

Bargiełówna, M. 1950. Grupy fonemów spółgłoskowych współczesnej polszczyzny kulturalnej. *Biuletyn Polskiego Towarzystwa Językoznawczego* 10: 1-25.

Bethin, C. 1992. *Polish syllables: the role of prosody in phonology and morphology*. Columbus: Ohio.

[19] Bailey and Hahn (2000) propose such a measure which involves complex mathematical calculations. For reasons of length limitations, we cannot discuss this proposal in this paper. The most popular phonotactic measure is the frequency of bigram occurrence and, less frequently, trigram frequency.

Coleman, J.S., and J. Pierrehumbert 1997. Stochastic phonological grammars and acceptability. *Computational Phonology* 3: 49-56.
Copeland, D.E., and G.A. Radvansky 2001. Phonological similarity in working memory. *Memory and Cognition* 29(5): 774-776.
Chomsky, N., and M. Halle 1968. *The sound pattern of English*. New York: Harper & Row.
Cyran, E., and E. Gussmann 1999 Consonantal clusters and governing relations: Polish initial consonant sequences. In H. van der Hulst and N. Ritter (eds.), *The syllable*. *Views and facts*, 219-247. Berlin: Mouton de Gruyter.
Dobrogowska, K. 1992. Word initial and word final consonant clusters in Polish popular science texts and in artistic prose. *Studia Phonetica Posnaniensia* 2: 47-121.
Dunaj, B. 1985. Grupy spółgłoskowe współczesnej polszczyzny mówionej (w języku mieszkańców Krakowa). *Zeszyty Naukowe UJ, Prace Językoznawcze 85*, Kraków.
Dziubalska-Kołaczyk, K. 2014. Explaining phonotactics using NAD. *Language Sciences* 46A: 6-17.
Frisch, S.A., N.R. Large and D.B. Pisoni 2000. Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language* 42(4): 481-496.
Grainger, J., H. Muneaux, F. Fabioli and J.C. Ziegler 2005. Effects of phonological and orthographic neigbourhood density interact in visual word recognition. *The Quarterly Journal of Experimental Psychology* 58A(6): 981-998.
Greenberg, J.H., and J.J. Jenkins 1964. Studies in the psychological correlates of the sound system of American English. *Word* 20: 157-177.
Hammond, M. 1999. *The phonology of English: a prosodic Optimality-Theoretic approach*. Oxford: Oxford University Press.
Hayes, B., and C. Wilson 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39: 379-440.
Jaskuła, K., and J. Szpyra-Kozłowska in press. *Wychódźc, Pcim* i *Rzgów.* Grupy spółgłoskowe w nazwach miejscowości w świetle fonotaktyki polskiej.
Jusczyk, P.W. 1997. *The discovery of spoken language*. Cambridge, Mass.: MIT Press.
Kreja, B. 1969. Z morfonologii i morfotaktyki współczesnej polszczyzny. *Prace Językoznawcze* 113. Wrocław: Ossolineum.
Kuryłowicz, J. 1952. Uwagi o polskich grupach spółgłoskowych. *Biuletyn Polskiego Towarzystwa Językoznawczego* 12: 221-232.
Laskowski, R. 1985. Z fonotaktyki polskich grup spółgłoskowych. *Studia Gramatyczne* 4: 35-52.
Leszczyński, Z. 1969. *Studia nad polskimi grupami spółgłoskowymi.* Wrocław: Ossolineum.
Metsala, J.L. 1997. An examination of word frequency and neighborhood density in the development of spoken word recognition. *Memory and Cognition* 25: 47-56.
Rochoń, M. 2000. *Optimality in complexity: the case of Polish consonant clusters.* Berlin: Akademie Verlag.

Rocławski, B. 1981. *System fonostatystyczny współczesnego języka polskiego*. Wrocław: Ossolineum.
Rubach, J., and G. Booij 1990. Edge of constituents effects in Polish. *Natural Language and Linguistic Theory* 8: 427-463.
Scarborough, R. 2012. Lexical similarity and speech production. Neighbourhoods for nonwords. *Lingua* 122(2): 164-176.
Storkel, H.L., J. Maekawa and J.R. Hoover 2010. Differentiating the effects of phonotactic probability and neighborhood density on vocabulary comprehension and production: a comparison of preschool children with versus without phonological delays. *Journal of Speech Language and Hearing Research* 53: 933-949.
Storkel, H.L. 2013. A corpus of consonant-vowel-consonant real words and nonwords: Comparison of phonotactic probability, neighbourhood density, and consonant age of acquisition. *Behaviour Research Methods* 45(4): 1159-1167.
Szpyra, J. 1995. *Three tiers in Polish and English phonology.* Lublin: Wydawnictwo UMCS.
Szpyra-Kozłowska, J., and P. Zydorowicz in press. Polish two-consonant clusters. A study in native speakers' phonotactic intuitions.
Śledziński, D. 2010. Analiza struktury grup spółgłoskowych w nagłosie oraz w wygłosie wyrazów w języku polskim. *Kwartalnik Językoznawczy* 3-4: 61-84.
Trask, R.L. 1996. *A dictionary of phonetics and phonology*. London and New York: Routledge.
Treiman, R. 1988. Distributional constraints and syllable structure in English. *Journal of Phonetics* 16: 221-229.
Ułaszyn, H. 1956. *Ze studiów nad grupami spółgłoskowymi w języku polskim*. Wrocław: Ossolineum.
Vitevitch, M.S. 2002. The influence of phonological similarity neighbourhoods on speech production. *Journal of Experimental Psychology, Learning, Memory and Cognition* 28(4): 735-747.
Vitevitch, M.S., and P.A. Luce 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40: 374-408.
Vitevitch, M.S., and P.A. Luce 2004. A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments and Computers* 36(3): 481-487.
Vitevitch, M.S., P.A. Luce, J. Charles-Luce and D. Kemmerer 1997. Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech* 40: 47-62.
Vitz, P.C., and B.S. Winkler 1973. Predicting the judged 'similarity of sound' of English words. *Journal of Verbal Language and Language Behaviour* 12: 373-388.
Zydorowicz, P., P. Orzechowska, M. Jankowski, K. Dziubalska-Kołaczyk, P. Wierzchoń, D. Pietrala 2016 *Phonotactics and morphonotactics of Polish and English. Theory, description, tools and applications*. Poznań: Wydawnictwo UAM.

Appendix 1.

Polish consonants:
Plosives: bilabial /p,b/, dental /t,d/, velar /k,g/
Affricates: dental /ts,dz/, alveolar /tʂ,dʐ/, prepalatal /tɕ, dʑ/
Fricatives: labio-dental /f,v/, dental /s,z/, alveolar /ʂ,ʐ/, prepalatal /ɕ,ʑ/, velar /x/
Nasals: bilabial /m/, dental /n/, prepalatal /ɲ/
Laterals: alveolar /l/
Trill: alveolar /r/
Semivowels: labio-velar /w/, palatal /j/
Before /i/ and /j/ nonpalatal consonants are palatalized and transcribed with /ʲ/.

Appendix 2. A list of the experimental items:

przun [pʂun], *szobl* [ʂɔbl], *mnep* [mnɛp], *luść* [luɕtɕ], *człać* [tʂwatɕ], *ruń* [ruɲ], *dzbyw* [dzbɨf], *wikr* [vʲkr], *gedź* [gɛtɕ], *skorz* [skɔʂ], *żaft* [ʐaft], *ciusk* [tɕusk], *szruń* [ʂruɲ], *żmyg* [ʐmɨk], *łetsz* [wɛtʂ], *wypt* [vɨpt], *kleg* [klɛk], *żlorz* [ʐlɔʂ], *siap* [ɕap], *dnysz* [dnɨʂ], *cekl* [tsɛkl], *spyp* [spɨp], *julm* [julm], *stryw* [strɨf], *tuśń* [tuɕɲ], *dbeś* [dbɛɕ], *siamn* [ɕamn], *chluf* [xluf], *cior* [tɕɔr], *dżyml* [dʑɨml], *słemf* [swɛmf], *dżacht* [dʑaxt], *rdzup* [rdzup], *kastrz* [kastʂ], *preń* [prɛɲ], *dzylsz* [dzɨlʂ], *lżag* [lʐak], *somf* [sɔmf], *lik* [lʲk], *styś* [stɨɕ], *rudźm* [rudʑm], *tkuf* [tkuf], *forst* [fɔrst], *gechć* [gɛxtɕ], *bolp* [bɔlp], *ćmesz* [tɕmɛʂ], *wulc* [vults], *głal* [gwal], *kefl* [kɛfl], *guf* [guf], *drecz* [drɛtʂ], *nidm* [ɲidm], *ciopń* [tɕɔpɲ], *zben* [zbɛn], *kształ* [kʂtaf], *binf* [bʲinf], *czwyp* [tʂfɨp], *ziaszcz* [ʑaʂtʂ], *mloń* [mlɔɲ], *szukt* [ʂukt], *fyks* [fɨks], *ciak* [tɕak], *rwol* [rvɔl], *gełc* [gɛwts], *szojf* [ʂɔjf], *łzur* [wzur], *sełsz* [sɛwʂ], *chliń* [xlʲiɲ], *kajcz* [kajtʂ], *kostw* [kɔstf], *ftać* [ftatɕ], *żems* [ʐɛms], *wziaj* [vʑaj], *szorl* [ʂɔrl], *wniup* [vɲup], *śpym* [ɕpɨm], *rojp* [rɔjp], *trzeg* [tʂɛk], *fupr* [fupr].