Are frequent, early and easy clusters also unmarked?

Katarzyna Dziubalska-Kołaczyk

Phonological theories generally assume that unmarked clusters would demonstrate similar behaviour in language use, language acquisition and language production. In particular, they are expected to be more frequent, acquired earlier and easier to produce than the marked ones. It seems, however, that the frequent, early and easy clusters are nonidentical sets. Additionally, morphonotactic clusters, which are by default more prone to be marked, behave differently than the marked phonotactic clusters, in particular, they are acquired earlier. In the paper, I will draw from the recent research by myself and colleagues[1] in order to discuss the difficulty in avoiding circularity in characterizing markedness as well as the omnipresent methodological bias connected with the choice of data, both experimental and collected, used to support claims concerning phonotactics.

1. Introduction

Five most frequent initial CC clusters in Polish are /pʃ, pr, st, mj, sp/ (cf. Zydorowicz *et al.* in press). Some studies (Dziubalska-Kołaczyk 1999, Zydorowicz 2010) on first language acquisition of Polish mention /sp, fʃ, vw, st, sw/ among those acquired early, and the morphonotactic[2] among them as acquired prior to the lexical, i.e. intramorphemic, ones (e.g. lexical /fʃ/ would be reduced much more often than the morphonotactic one). Assessment of all those clusters from the perspective of speech production is a complex task, involving detailed phonetic description focused on coarticulatory effects and gestural coordination.[3] Importantly, production-based criteria interact with the acoustic/auditory ones. Additionally, such descriptions are always theory-grounded and therefore provide divergent assessments of 'ease'. Already a shift from a production to a perception perspective leads to conflicting interpretations. For example, a homorganic cluster is 'easier' to produce but 'less easy' to perceive. Despite the above problems of analysis, what can be noticed is that /st/ and /sp/ clusters figure in both of the above sets of the frequent and of the early acquired clusters. It is well known that s+stop clusters are notoriously unaccountable for across models of phonotactic markedness. Sonority-based models generally fail to explain their occurrence. Phonetic, phonological and typological accounts point to the uniqueness of those clusters (cf. an overview by Olender 2013). Still, phonologically speaking, they are marked, and require a special treatment across models. As already mentioned, they are also frequent and early, possibly also easy.

The above example introduces the topic of this paper: the relationship between independently defined markedness and frequency, order of acquisition and ease of production. The major advantage of the concept of markedness has been to allow for predictions of the occurrence of structures in given types of linguistic behaviour. The less marked (=more natural) a structure, the more likely it would be to occur frequently, to be acquired early and to be easy to produce. Of course, such inference may only work on condition that markedness be defined independently of the criteria of frequency, age of acquisition or ease of articulation. Otherwise, it becomes circular.

In this paper we deal with markedness in phonotactics. Clusters are deemed more or less natural depending on a measure called NAD (Net Auditory Distance). One expects the clusters preferred by NAD to be more frequent, to be acquired before the dispreferred ones as well as to be easier to produce. First, the phonological theory underlying the model of phonotactics will be briefly introduced (section 2) and NAD as a measure of markedness will be presented (section 3). Next, the data illustrating frequency, acquisition and ease will be discussed and assessed from the perspective of NAD (section 4). Section 5 will contain conclusions.

2. Beats-and-Binding (B&B) phonotactics

The model of phonotactics operating with NAD (cf. Dziubalska-Kołaczyk 2009, 2014) is part of the syllable-less theory of phonology called Beats-and-Binding (B&B) phonology (Dziubalska-Kołaczyk 2002). The rationale behind this model of phonotactics is to counteract the preference for CV (but cf. Hyman 2008), which overwhelmingly forces clusters to wither either via reduction or epenthesis/prothesis or at least substitution (CCVCV, CCV→CVCV, CCV→VCCV). Clusters of consonants persist thanks to auditory contrast between members of a cluster and its satisfactory distribution across the word. The principle of perceptual contrast is grounded in the semiotic figure-and-ground principle, on the one hand, and in phonetics, on the other, since as observed by Ohala (1990), larger modulations in several acoustic parameters (amplitude, periodicity, spectral shape, F0) have more survival value than lesser ones and therefore will persist in languages.

3. NAD

3.1. Definition of NAD

The Net Auditory Distance (NAD) Principle defines cluster preferability in relation to the position in the word (initial, medial and final). It reads:

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

For example, for the initial cluster C1C2V:

NAD (C1,C2) ≥ NAD (C2,V). This notation means that, in word-initial double clusters, NAD between the two consonants should be greater than or equal to NAD between a vowel and a consonant neighbouring on it.

3.2. Measuring NAD

It is believed (Dziubalska-Kołaczyk 2002, 2009, 2014) that auditory (perceptual) distance can be expressed by combinations of articulatory features, which eventually bring about the auditory effect. An indefinite number of articulatory features as well as detailed acoustic cues would have to be investigated in terms of the degree of their contribution to the overall auditory effect obtained in a cluster. As a starting point, manner of articulation (MOA) and place of articulation (POA) are selected, as well as the S/O (sonorant/obstruent) distinction. Tables 1 and 2 illustrate the way in which number values are being assigned to particular POAs and MOAs. The presence of the S/O distinction is signalled by 1, its absence by 0.

The manners and places assumed in the tables are selected according to their potential relevance: 6 manners (stop, affricate, fricative, sonorant stop, approximant, semivowel), where affricates and semivowels are, tentatively, attributed half a distance due to their ambiguous nature; and 5 places (labial, coronal, dorsal, radical and laryngeal or glottal). Manners refer to the most generally acknowledged version of the so-called sonority scale, while places are taken from Ladefoged (2006: 258). Both lists are extendible and modifiable (e.g., Ladefoged's list consists of 5 nodes which branch into 12 more detailed features), depending on the amount of detail we want to include in the definition of distance. Importantly, vowels need to be differentiated according to their colour, too, which will allow reflecting the intersegmental distances with a higher degree of precision.

Table 1 is meant as a general, universal reference for the values of MOA and POA which constitutes a guideline for the creation of language-specific tables. Table 2 is the adaptation of Table 1 to Polish. Polish has three values for obstruents, two values for liquids, an additional distinction within glides (nasal vs. oral), has a value for labiodental within labials, three values within the feature coronal, and two values within the feature dorsal.

Table 1. Distances in MOA and POA.

4	3	2	1	0	
OBSTRUENT		SONORANT		VOWEL	
STOP	FRICATIVE	SONORANT STOP	APPROXIMANT		
(AFFRICATE)			semiV		
p b	ɸ β f v	m ɱ	w	labial	1
t d ʈ ɖ	θ ð s z ʂ ʐ ʃ ʒ	n	r l	coronal	2
k g c ɟ	ɕ ʑ x ɣ	ɲ ŋ	j	dorsal	3
				radical	4
ʔ	h			laryngeal (glottal)	5

Table 2. Distances in MOA and POA: Polish.

OBSTRUENT			SONORANT					VOWEL		
STOP		FRICATIVE	NASAL	LIQUID		GLIDE				
	AFFRICATE			2.0						
				lateral	rhotic					
5.0	4.5	4.0	3.0	2.5	2.0	1.5	1.0	0		
p b			m			w̃	w	1.0	bilabial	LABIAL
		f v						1.5	labio-dental	
t d ʦ ʣ s z			n	l				2.0	(post-) dental	CORONAL
tʂ dʐ ʂ ʐ				r				2.3	alveolar	
ʨ ʥ ɕ ʑ			ɲ					2.6	alveolo-palatal	
						j̃	j	3.0	palatal	DORSAL
k g		x	ŋ			w̃	w	3.5	velar	
								4.0		RADICAL
								5.0		GLOTTAL

3.3. An example of a calculation[4]

For an initial Polish cluster /prV/, NAD preferences should be calculated as follows:

NAD CC = |MOA1 - MOA2| + |POA1 - POA2| + S/O
NAD CV = |MOA1 - MOA2| + S/O
NAD CC = |5-2|+|1-2.3|+1 = 3+1.3+1 = 5.3
NAD CV = |2-0|+0 = 2+0=2
5.3 > 2

So, the preference NAD (C1,C2) ≥ NAD (C2,V) is observed since 5.3 > 2.

3.4. NAD as a measure of markedness: predictions

NAD allows for the formulation of two types of predictions. On the one hand, it distinguishes between two sets of clusters, preferred (i.e. those which observe the preferences) and dispreferred ones (those which do not), which constitutes a binary qualification into unmarked=preferred=natural vs. marked=dispreferred. On the other hand, it allows to build a hierarchy of preference of clusters ranging from the most preferred to the most dispreferred (least preferred), which constitutes a scalar qualification of relative markedness from the most natural=unmarked=preferred to the heavily marked. This clearly is a strong attribute of this measure, since it reflects the human ability to both categorize and build continua. The matrices below (Figure 1, 2 and 3, prepared by Michał Jankowski) provide an

	p	b	t	d	k	g	ts	dz	tʃ	dʒ	tɕ	dʑ	f	v	s	z	ʃ	ʒ	ɕ	ʑ	x	n	ɲ	m	ŋ	l	r	w	j	w̃	j̃
p			NO						NO						NO		NO										NO				
b																															
t									NO				NO				NO							NO		NO	NO	NO			
d																												NO			
k			NO												NO						NO					NO		NO			
g																												NO			
ts																															
dz																															
tʃ	NO		NO																												
dʒ																															
tɕ																					NO										
dʑ																															
f			NO				NO				NO																NO				
v																															
s	NO		NO		NO								NO											NO				NO			
z																						NO		NO				NO			
ʃ	NO		NO						NO				NO								NO										
ʒ																															
ɕ	NO										NO												NO	NO		NO					
ʑ																							NO								
x			NO																NO									NO			
n			YES	NO			NO		NO						NO		NO					NO									
ɲ							YES		NO		NO	NO																			
m	YES												NO				YES		YES			NO									
ŋ					YES																										
l			YES		YES		YES		YES								YES						NO	NO							
r	YES		YES		YES		YES		YES		YES		YES		YES		YES		YES			NO	NO	YES				NO			
w	YES		YES		YES						YES		YES										YES	YES				NO			
j			YES		YES						YES						YES		YES		YES			YES							
w̃	YES														YES		YES				YES							NO			
j̃																			YES	YES											

Figure 1. A dispersion matrix for initial CC clusters according to NAD.

	p	b	t	d	k	g	ts	dz	tʃ	dʒ	tɕ	dʑ	f	v	s	z	ʃ	ʒ	ɕ	ʑ	x	n	ɲ	m	ŋ	l	r	w	j	w̃	j̃
p			NO												NO		NO		NO		NO		YES			YES	YES	YES	YES		
b																NO		NO		NO						YES	YES	YES	YES		
t					NO								NO				NO				NO	YES	YES			YES	YES	YES			
d		NO												NO				NO				NO	NO	YES		YES	YES	YES	YES		
k			NO										NO				NO		NO							YES	YES	YES	YES		
g				NO								NO		NO				NO					NO	YES		YES	YES	YES	YES		
ts																								YES		YES		YES			
dz		NO												NO																	
tʃ									NO		NO		NO															YES			
dʒ										NO																					
tɕ													NO											YES							
dʑ														NO																	
f	NO		NO				NO		NO		NO				NO		NO		NO		NO					YES	YES		YES		
v				NO								NO				NO		NO		NO		NO	NO	NO		NO	YES	YES	YES		
s	NO		NO		NO		NO						NO		NO		NO		NO		NO	NO		YES			YES	YES			
z		NO		NO		NO						NO		NO				NO				NO	NO	NO		NO	YES	YES	YES		
ʃ	NO		NO		NO				NO				NO									NO		YES		YES	YES	YES			
ʒ																							NO	NO							
ɕ	NO										NO		NO										NO	YES		YES	YES				
ʑ																										NO	YES				
x							NO				NO		NO				NO							YES		YES	YES	YES			
n																															
ɲ																															
m																					NO	NO				NO	YES	YES	YES		
ŋ																															
l														NO				NO				NO	NO								
r								NO						NO																	
w		NO														NO				NO											
j																															
w̃																															
j̃																															

Figure 2. A dispersion matrix for final CC clusters according to NAD.

	p	b	t	d	k	g	ts	dz	tʃ	dʒ	tɕ	dʑ	f	v	s	z	ʃ	ʒ	ɕ	ʑ	x	n	ɲ	m	ŋ	l	r	w	j	w̃	j̃
p			YES		YES		YES		YES		YES		YES		YES		YES		YES		YES	NO	NO	NO		NO	NO	NO	NO		
b																		YES				NO	NO	YES		NO	NO	NO	NO		
t	YES		YES		YES		YES		YES		YES		YES		YES		YES		YES		YES	NO	NO	NO		NO	NO	NO	NO		
d		YES		YES								YES		YES				YES				YES	YES	NO		NO	NO	NO	NO		
k		YES	YES		YES		YES				YES		YES		YES		YES		YES			NO	NO			NO	NO	NO	NO		
g				YES								YES				YES		YES				NO	YES	NO		NO	NO	NO	NO		
ts			YES		YES		YES				YES											YES	NO			NO			NO		
dz														YES				YES				YES		YES							
tʃ	YES		YES		YES		YES		YES		YES						YES					YES	YES	NO		NO					
dʒ		YES																													
tɕ		YES					YES				YES		YES				YES							NO							
dʑ																		YES						NO							
f			YES		YES		YES		YES		YES				YES		YES		YES			YES	NO	YES		NO	NO		NO		
v				YES				YES				YES				YES		YES				YES	YES	YES		YES	NO		NO		
s	YES		YES		YES		YES		YES		YES		YES		YES		YES				YES	YES		NO			NO	NO	NO		
z		YES		YES		YES						YES		YES				YES				YES	YES	YES		YES	NO	NO	NO		
ʃ	YES		YES		YES		YES		YES		YES		YES		YES		YES		YES		YES	YES	YES	NO		NO		NO			
ʒ		YES		YES						YES				YES				YES				YES	YES	YES		YES					
ɕ		YES									YES						YES					YES	YES	NO		NO	NO				
ʑ		YES										YES		YES				YES				YES	YES	YES		YES	NO				
x			YES		YES		YES		YES		YES		YES									NO	YES	NO		NO	NO	NO			
n			YES	YES	NO		YES	YES	YES		NO	YES	YES	YES	YES	YES		YES	YES			YES	YES								
ɲ					NO		NO		YES		YES	YES						YES						YES							
m	YES	YES	NO		NO		NO		NO		NO		YES	YES			NO	YES				YES	YES	YES			NO		NO		
ŋ					YES	YES																									
l		NO	NO	YES	NO	NO	NO	YES	NO		NO	NO	NO	YES			NO	YES	NO		NO	YES	YES	YES		YES			NO		
r	NO	NO	NO	NO	NO	NO	NO	NO	NO		NO	NO	NO	NO	NO		NO	YES	NO		NO	YES	YES	NO		YES	YES	YES	YES		
w	NO	NO	NO	NO	NO	NO	NO		NO		NO	NO	NO	NO		NO	NO	NO				NO	NO	NO		YES	YES	YES			
j	NO	NO	NO	NO	NO	NO	NO		NO		NO	NO		NO	NO	NO	NO	NO	NO	NO		NO	NO	NO		NO	YES				
w̃	NO	NO													NO	NO		NO			NO					NO		YES			
j̃																			NO	NO											

Figure 3. A dispersion matrix for medial CC clusters according to NAD.

adequate visualisation of the nature of NAD predictions. The matrices have been derived for the CC pairs in all word-contexts (initial, medial, final) for 31 Polish consonants (ordered by manner/place in columns/rows). They demonstrate (a) the non-overlapping sets of clusters in initial, medial and final position (which testifies to the logical integrity of the preference formulae: a given cluster cannot be equally good in different word positions), (b) the distinct sets of preferred and dispreferred clusters (YES's and NO's), and (c) the continua of preference across the matrices from the most to the least preferred cluster (e.g., in Figure 1, pj is the most preferred cluster, followed by pw and bj (with the same value of NAD), followed by pr and bw (ditto), followed by pl, br, tw and dj (ditto), etc.). The key attribute of the matrices is that they reflect simultaneously the dichotomic (categorical) and scalar (gradual) evaluation of clusters. Human perception is both categorical and gradual, too.

4. NAD and the data

4.1. NAD and frequency

The consonant sequences were extracted from three types of corpora of Polish: a list of lemmas based on a 8 thousand dictionary of 'core' Polish ('dictionary' henceforth), a list of inflectional forms generated on the basis of standard paradigms for Polish nouns, adjectives and verbs ('paradigms' henceforth), and a list of over 500,000 inflectional forms derived from a collection of newspaper texts ('corpus', henceforth). The result was a list of approximately 1,500 consonant cluster types of all sizes and in all positions.

Nearly half of the NAD-preferred 60 initial CC clusters from the Polish matrix (cf. Figure 1 above) were compared to first 5 most frequent Polish initial CC clusters in the 'corpus', the 'paradigms' and the 'dictionary' and first 14 most frequent Polish initial CC clusters in the 'dictionary' and the 'corpus'. We noted a difference between 'corpus' and 'dictionary' in types of clusters and a difference between the sets of 5 and 14 clusters in the percentage of unmarked clusters. Table 3 below summarizes the results. Most importantly, there is an essential, qualitative difference in what we conclude depending on how much of the data we consider. If we look arbitrarily at the first 5 most frequent clusters in the language, 3 out of 5 are marked, which points to NAD as a weak predictor of frequency of occurrence. If, however, we examine a larger set (in this case 14 most frequent clusters were considered due to a notable drop in frequency of further clusters), it turns outs that 50% of the clusters in the 'paradigms' and 'corpus' and 71% of the clusters in the 'dictionary' are NAD-preferred, i.e. relatively unmarked. Therefore, NAD scores as a good predictor for the 'dictionary' data only.

Table 3. Initial CC clusters: NAD and frequencies.

preferred according to NAD									frequent in ‘paradigms’ & ‘corpus’	frequent in ‘dictionary’
5	4.25	4	3.5	3.3	3.25-3.2	3	2.5	2.3-2	pʃ, pr, st mj, sp	pʃ, pr, st, sp, pj
pj	kw *pw*	bj	fj *kj*	pr	bw gw *kr* tw xw	dj	gj *kl* tʃw tsw *vj*	*tr* *sw* *gr* mj pl zj	dl, *vj*, kt. gd, *pj*, *kj*. zn, *kr*, *tr*	*vj*, *tr*, *kr*, *sw*, *gr*, sk, *pw*, *kl*, *mj*
preferred according to NAD in 5 most frequent clusters (underline) *preferred according to NAD in 14 most frequent clusters (italic)*										

Next, if we examine the marked clusters from among the most frequent 14 in the ‘corpus’, we notice a huge discrepancy between the ‘dictionary’ and the ‘corpus’ frequencies (roughly reflecting the difference between type and token frequencies, respectively; cf. Table 4). This is accounted for by usage: the functional load of words containing many of these marked clusters is extremely high in Polish.

Table 4. Initial CC clusters: type vs. token frequency.

cluster	dictionary	corpus	words
/dl /	3	172698	*dla* ‘for’
/kt/	7	92997	*kto* ‘who’, *który* ‘which’
/gd/	3	75041	*gdy* ‘when’
/zn/	19	35424	*znać* ‘know’

Earlier in Table 3 we saw the failure of NAD to predict the frequent clusters in the ‘corpus’ (50% score). Data in Table 4 shows that 4 out of 7 dispreferred clusters (i.e. another 29% of all 14) have very low type frequency, which gives credit to NAD that predicts them to be dispreferred. In other words, in this case NAD relates well with type frequency and not at all with token frequency.

4.2. NAD: criteria manipulation

NAD calculations may show different results depending on the choice and use of criteria. We have already mentioned scalar vs. binary qualification of the ‘goodness’ of clusters (cf. section 3.4. above). The matrices (Figures 1-3 above) show both binary and scalar typology of clusters. The analysis of frequencies demonstrates the advantage of scalar qualification since it reflects *relative* markedness, predicted by the *preference* theory (cf. Dressler 1999). For instance, in Table 3 above the "preferred according to NAD" clusters are distributed along 9 values within the category *preferred*.

The set of features used for the calculations may be changed or expanded. For example, laryngeal features are non-redundant within obstruents and largely redundant within sonorants. Therefore, instead of Lx (difference in the value of voicing),[5] the S/O criterion, i.e. the difference between sonorant and obstruent (set at 1) has been introduced into the equation (cf. section 3.2 above). It makes a difference for (calculation formulae as in section 3.3 above):

–initial clusters of voiced obstruents with sonorants (in this order; they improve), e.g.
/vj/ 2.5 (Lx contributes 0) 3.5 (S/O contributes 1)
/gr/ 2.2 (Lx contributes 0) 3.2 (S/O contributes 1)
–initial clusters of voiced obstruents (they worsen), e.g.
/zd/ -4 (Lx contributes 0) -5 (S/O contributes |-1|)

More detailed features of MOA and POA could be introduced and their values could be modified, for example by introducing weights (e.g. Calderone & Bertinetto 2013 experimented with the proposal of 1.0 for CV opposition, 0,8 for manners and 0.5 for places and voicing as input values to their probabilistic system). Language-specific values can be introduced (cf. NAD for Polish vs. English) in order to reflect different organization of segmental inventories across languages, entailing different degrees of exploitation of a given feature. POA's for vowels need to be included, too. Additionally, more phonetic detail could be considered when specifying the positional preferences, e.g. the timing differences between initial and final clusters.[6]

All the modifications aim at setting the values so that the calculations maximally match the data. In this way the model (deductive) receives feedback (inductive) from the data.

4.3. NAD and acquisition: collected data

The order of acquisition of different types of clusters appears to vary across languages. For example, Fikkert & Freitas (2004) showed the following sequences of acquisition of clusters by Dutch and European Portuguese children:

PL > FL > SP Dutch children
SP > PL > FL European Portuguese children

Thus, SP (sibilant-plosive) clusters were acquired last by the Dutch children and first by the European Portuguese kids. Additionally, some Dutch children had SP first. Frequency effect was ruled out since in both languages CL (obstruent-liquid) clusters (PL plosive-liquid and FL fricative-liquid) are much more frequent than SP (10% vs. 3% of the kids intake). Also, CL emerged long before SP (and SP long before CL respectively) in both languages (at the 3-6 months span). As we will see, for Polish children none of this is true.

4.4. NAD and acquisition: experimental data

A picture naming task was used to elicit the production of words with six initial clusters /st, sp, sk, sx, sm, sw/ by 49 Polish kindergarten children, half of them with normal phonological development, mean age of 42 months, the other half with DPD (developmental phonological disorder), mean age of 57 months. The children were selected out of 114 according to the results of a battery of tests testing their overall performance, among others their ability to produce singleton consonants included in the studied clusters (Yavaş and Marecka 2013). The extent of reductions from the most reduced to the best preserved was as follows:

sx > sw > *sk* > *sp* > sm > *st*

As can be seen, both ends of the scale as well as the middle are occupied by marked (=dispreferred) clusters (signalled by italics). The group of children with DPD reduced the majority of the productions, but showed similar tendencies (cf. also Marecka & Dziubalska-Kołaczyk 2014).

However, the experiments involved an arbitrary selection of the data for elicitation, without reference to the systems of the languages studied. For instance, for word initial *st* in Polish an input frequency effect may be considered, since this is the third most frequent cluster in Polish (both type and token frequency). As Fikkert & Freitas (2004: 10) observed "it is important to consider the language system as a whole to interpret the data, both to explain differences between children acquiring the same language (i.e. the child's own phonological system determines what optimal realizations for clusters are), and between children acquiring different languages".

4.5. NAD, frequency and acquisition

Table 5 presents once again the data of Table 3, but with the addition of data coming from one case study, namely, the analysis of clusters produced by Zosia, a young girl from Poznan, who was recorded between 1;7 and 3;2 years of age according to a longitudinal production study (cf. Zydorowicz 2010). Marked clusters that are frequent in Zosia's speech (shown as strikethrough) can be explained by a strong frequency effect, i.e. by statistical learning. These acquisitional data furthermore confirm that small sets of data (e.g. 5 early clusters) are not sufficient to draw conclusions about the predictive power of NAD, whereas a larger pool of data demonstrates that:

– the majority of frequent clusters is unmarked in the 'dictionary' (78%)
– approximately 50% of frequent clusters are unmarked in the 'corpus'
– approximately 30% of first acquired clusters are unmarked.

The top cluster (both as type and token frequency) is the same in the most frequent list and in the first acquired list; this cluster is /pʃ/, and it is a marked structure according to NAD. Its occurrence can be explained by morphological productivity since /pʃ/ appears within highly productive affixes, e.g. in *przed-*, *przy-*, *prze-*, all beginning with /pʃ/.

Table 5. Initial CC clusters: NAD, frequencies and L1 acquisition.

preferred according to NAD									Frequent in 'paradigm' & 'corpus'	Frequent in 'dictionary'	L1 acquisition by Zosia 2.8-2.9
5	4.2 -5	4	3.5	3.3	3.25 -3.2	3	2.5	2.3 -2	~~pʃ~~. pr, ~~st~~, mj. ~~sp~~ dl, vj. kt. gd, pj, kj. zn, kr. tr	~~pʃ~~. pr, ~~st, sp.~~ pj vj. tr. kr. sw, gr. ~~sk~~, pw, kl. mj	(33) pʃ, (12) st. (9) kc, (7) mʃ, (6) cp, (5) dv. kl, xts, (4) ~~sp~~, tʃt, sw, xw, (3) pw. gw, ~~sk~~, mn
pj	kw pw	bj	fj kj	pr	bw gw kr tw xw	dj	gj kl tʃw tsw vj	tr sw gr mj pl zj			
preferred according to NAD in 5 most frequent clusters (underline) *preferred according to* NAD in 14 most frequent clusters and L1 (italic) ~~overlapping marked in frequent clusters and L1 (strikethrough)~~ The numbers in parentheses in the last column of Table 5 are tokens in Zosia's speech.											

4.6. NAD, cluster size and ease

In the above sections we discussed frequent and early clusters and how they score in terms of markedness. Let us consider 'easy clusters' in this perspective, too. Of the two measures of markedness of clusters used in this study, NAD and the size of a cluster, the latter is relatively straightforward with respect to ease: a shorter cluster is easier to articulate and to hear than a longer one. NAD's predictions of ease are no longer so straightforward since NAD is a derivative of articulatory and acoustic features contributing in total to the final auditory effect. Characterising ease from this perspective is ambiguous. We do not know, for instance, what is easier: sustaining a gesture or changing (=modulation of) the gesture (cf. e.g. /wu/ or /ji/ in which a labial or palatal gesture is sustained while the perception of those clusters is very poor). There is an apparent conflict between perceptual contrast and ease of articulation; however, the auditory distance obtained within a cluster is the result of an optimal compromise between the two. In casual speech, clusters reduce and change, and the modifications are much more often due to assimilation than dissimilation (i.e., due to the tendency towards ease of articulation). Still, although the general tendency is to reduce clusters, marked (=dispreferred) clusters arise, too (cf. Dressler *et al.* 2001), so the tendency towards ease may result in the actual markedness.

To provide an illustration of the relationship between markedness and ease, let's consider long clusters, first. We classified 4, 5 and 6 consonant clusters as long (cf. Table 6). We analysed 273 long clusters, which constituted 15% of all clusters of Polish (cf. section 4.1. above). The low percentage of the types (for example, only 3 clusters of 6 consonants) implies that the size of a cluster does predict difficulty of production and, consequently, of perception (i.e., long clusters are difficult and thus rare). Additionally, almost 100% of those clusters are morphonotactic (see Table 7). Morphonotactic clusters are expected to be marked since in this way they signal morphological boundaries (cf. Dressler & Dziubalska-Kołaczyk 2006). In the case of morphonotactic long clusters, however, there is the additional aspect of size resulting from the concatenation of morphemes that already contain clusters. Such sizeable clusters are bound to be difficult to pronounce. Markedness expressed by size is thus a good predictor of phonetic difficulty .

Table 6. Cluster length in Polish.

number of C	number of types
2	485
3	976
4	219
5	51
6	3
	1734

Table 7. Long clusters: phonotactics vs. morphonotactics.

Cluster length (in number of phonemes)	Lexical (intra-morphemic)	Morphono-tactic	Examples
6	–	100%	przest<u>ępstwi</u>e, wewn<u>ątrzwsp</u>ólnotowy, wewn<u>ątrzzwią</u>zkowy
5	–	100%	dzieci<u>ństwi</u>e, be<u>zwzgl</u>ędnie, ro<u>zbrzmi</u>ewa, ko<u>ntrprz</u>ykład, po<u>stzwią</u>zkowy, tysi<u>ącstr</u>onicowy
4	5%	95%	/rstf/ wa<u>rstw</u>a (lex 3%) ministe<u>rstw</u>o (-stwo morph 97%), <u>krwią</u> (lex), krn<u>ąbrn</u>y (lex), <u>źdźbł</u>o (lex), ro<u>ztkl</u>iwiać (morph), o<u>dstrz</u>ał (morph), j<u>ędrn</u>y (lex)

Markedness measured by NAD, however, may fail as a predictor of difficulty in specific types of performance, in particular, when marked (NAD dispreferred) clusters arise in contexts triggered by the principle of least effort. The application of phonostylistic processes in casual speech, such as vowel reduction "may lead to a temporary rise of 'new' clusters which are marked" (Dressler *et al.* 2001: 108). In the casual, everyday style, we found the following 'new' clusters in Polish (cf. Dressler *et al.* 2001: 108, Table 8), all of them marked. In this case, the natural speaker-friendly tendency for ease led speakers to produce long clusters.

Table 8. "New" clusters resulting from vowel reduction.

Double clusters (12)	nč	n̩ č	ʒč	m̩ m	s̩ s	ss(2)	m̩ v	zdz	zz	z̩ ts	n̩ d
Triple clusters (9)	tlk	t l̩ k	pʃk	fʃs	m̩ vj	pʃn	ssn	pʃp	s̩ td		
Four-consonant clusters (4)	fʃsk	pkʃw	pʃʃw (2)								
Five-consonant clusters (2)	fʃstk	fʃsts									

5. Conclusions

We have shown a better implicational relation between the unmarked and frequent clusters than between the unmarked and first acquired ones. Additionally, unmarked implies frequent in a dictionary better than in a corpus. Generally, markedness predictions based on NAD appear to work better for the system (lexicon) and its use (dictionary and corpus) than for its acquisition. It may be that we deal with a critical time in first language acquisition when the markedness predictions begin to operate. This, however, remains to be investigated. Markedness predictions of ease of production have been only shortly discussed in the paper and shown to depend on the applied measure of markedness (size of cluster vs. NAD) as well as the type of performance.

We conclude that for markedness to be a reliable tool of analysis we need to consider the whole system of the language studied, cross-compare to other systems, discover language-specific constraints, collect and elicit larger quantities of reliable, natural, representative and detailed data, as well as discover more criteria which co-determine markedness.

Address of the Author

Katarzyna Dziubalska-Kołaczyk Faculty of English, Adam Mickiewicz University in Poznań, al. Niepodległości 4, 61-874 Poznań, Poland <dka-sia@wa.amu.edu.pl≥

Notes

[1] Michał Jankowski, Paulina Zydorowicz.
[2] Morphonotactics is the area of interaction between morphotactics and phonotactics (cf. Dressler and Dziubalska-Kołaczyk 2006).
[3] As opposed to a purely phonemic treatment of clusters, reflected in their phonemic transcription.
[4] For the purposes of the calculations, the phonotactic calculator has been developed (consecutive versions by Krynicki, Pietrala, the current one by Aperliński), available online under http://wa.amu.edu.pl/nadcalc/.
[5] We included the voicing contrast, symbolized by Lx, in the previous version of NAD. The presence of the voicing contrast between two members of a cluster was signaled by the value 1, its absence by 0.
[6] "Onset clusters are considered to be timed more rigidly, allowing for less timing variability and being less overlapped than heterosyllabic or coda clusters (Byrd 1996; Hardcastle & Roach 1997)" (Bombien 2011: 35).

Bibliographical References

Baroni Antonio 2014. On the importance of being noticed: the role of acoustic salience in phonotactics (and casual speech). *Language Sciences* 46. 18-36.

Bombien Lasse 2011. *Segmental and prosodic aspects in the production of consonant clusters – On the goodness of clusters*. PhD dissertation, München.

Calderone Basilio & Pier Marco Bertinetto 2013. From Phonotactics to Syllables. A psycho-computational approach. Talk delivered during the 46th *Societas Linguistica Europaea*, 17-20 Sept, 2013, Split.

Dressler Wolfgang U. 1999. On a semiotic theory of preferences in language. *The Peirce Seminar Papers* vol. 4. New York: Bergham Books. 389-415.

Dressler Wolfgang U., Dziubalska-Kolaczyk Katarzyna & Rossella Spina 2001. Sources of markedness in language structures. *Folia Linguistica Historica* 22 (1-2). 103-135.

Dressler Wolfgang U. & Katarzyna Dziubalska-Kolaczyk 2006. Proposing morphonotactics. *Rivista di Linguistica* 18(2). 249-266.

Dziubalska-Kołaczyk Katarzyna 1999. Early L1 clusters and how they relate to universal phonotactic constraints. *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS)*, San Francisco. 317-320.

Dziubalska-Kołaczyk Katarzyna2002. *Beats-and-Binding Phonology*. Frankfurt/Main: Peter Lang.

Dziubalska-Kołaczyk Katarzyna 2009. NP extension: B&B phonotactics. *PSiCL* 45 (1). 55-71.

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

Fikkert Paula & Maria João Freitas 2004. The role of language-specific phonotactics in the acquisition of onset clusters. In L. Cornips & J. Doetjes (eds.), *Linguistics in the Netherlands* 2004. Amsterdam: John Benjamins. 1-12.

Hyman Larry 2008. Universals in phonology. *The Linguistic Review* 25 (1-2). 83-137.

Ladefoged Peter 2006. *A Course in Phonetics*. 5th edition. Boston: Heinle & Heinle.

Marecka Marta & Katarzyna Dziubalska-Kołaczyk 2014. Evaluating the models of phonotactic constraints on the basis of the sC cluster acquisition data. *Language Sciences* 46. 37-47.

Ohala John J. 1990. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. The Parasession on the Syllable in Phonetics and Phonology. *CLS* 26,2. 319-338.

Olender Adam 2013. *A synchronic investigation of leftmost /s/ + stops*. M.A. dissertation, Adam Mickiewicz University in Poznań.

Pietrala Dawid 2014. *The application of linguistic principles in programming: phonemic transcription and phonotactic constraints*. Unpublished Ph.D. dissertation, Adam Mickiewicz University in Poznań.

Yavaş Mehmet & Marta Marecka 2013. Acquisition of Polish #sC clusters in typically-developing children and in children with phonological disorders. *International Journal of Speech-Language Pathology*.

Zydorowicz Paulina 2010. Consonant clusters across morpheme boundaries: Polish morphonotactic inventory and its acquisition. *PSiCL* 46(4). 565-588.

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