The NAD Phonotactic Calculator – an online tool to calculate cluster preference in English, Polish and other languages.

Katarzyna Dziubalska-Kołaczyk, Dawid Pietrala, Grzegorz Aperliński
NAD Phonotactic Calculator: http://wa.amu.edu.pl/nadcalc/

Background

The Beats-and-Binding Phonology (B&B henceforth) is a theory of phonology proposed by Dziubalska-Kołaczyk (1995, 2002a, 2009, 2010, 2011). B&B is a syllable-less model of phonology which is derived from the framework of Natural Linguistics and Natural Phonology[1]. The B&B phonotactics constitutes a crucial part of the entire model because it accounts for the organisation of consonant clusters in language. The model also aims at predicting the behaviour of consonant clusters in first and second language acquisition, speech disorders and diachronic change. The B&B model replaces traditional syllable components (onset, nucleus, coda) with *beats* and *non-beats*:

- Beats – best realized by vowels, relatively more prominent,
- Non-beats – best realized by consonants, relatively less prominent.

Vowels typically act as beats as they are more salient thanks to high sonority and articulatory openness, however, consonants may at times be beats as well (for instance, syllabic consonants in English). Beats and non-beats form relations called bindings. Bindings connect beats (B) and non-beats (nB) in a sequence. Typically, sequences of beats and non-beats alternate so as to avoid hiatuses, i.e. the contact of two beats. This is often achieved by by the insertion of consonants (non-beats) between the two beats.

Measuring cluster preferability

To calculate the NAD of two segments, the following formula is used:

MOA + POA[2]

[1] Stampe (1969, 1979), Donegan and Stampe (1979), Dressler (1985), Dziubalska-Kołaczyk (1995, 2001a, 2001b, 2002a, 2002b, 2002c, 2007), and Dressler and Dziubalska-Kołaczyk (2006).

[2] At the moment the sonorant/obstruent (S/O) contrast is being analyzed as an additional criterion that could be included in the formula (Dziubalska-Kołaczyk 2014, in press).

To illustrate the cluster evaluation with an example, let us analyze the NAD condition for initial doubles:

C_1C_2V	NAD (C1,C2) ≥ NAD (C2,V)

Fig. 1. The phonotactic preference condition for double initial clusters.

The condition reads:

In word-initial double clusters, the net auditory distance (NAD) between the two consonants should be greater than or equal to the net auditory distance between a vowel and a consonant neighbouring on it.

The NAD comprises manner and place of articulation, which notationally can be written as |MOA| + |POA|. For instance, in order to calculate NAD preferability for a Polish double initial cluster /pr/, it is necessary to find out the MOA and POA values for /p/ and /r/ as well as MOA for the vowel which follows the cluster (hereafter V[3]). The values can be read from tables provided by Dziubalska-Kołaczyk (2014: 10)[4]:

POLISH CONSONANTS

OBSTRUENT			**SONORANT**					**VOWEL**		
STOP		**FRICATIVE**	**NASAL**	**LIQUID**		**GLIDE**				
	AFFRICATE			lateral	rhotic					
5.0	4.5	4.0	3.0	2.5	2.0	1.5	1.0	0		
p b			m			w	w̃	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	
						j j̃		3.0	palatal	**DORSAL**
k g		x	ŋ			w w̃		3.5	velar	

[3] In the present formula all vowels have equal values for MOA, i.e. 0. POA is not yet included for vowels. In the future version of the calculator, POA for vowels will be translated into vowel colours.

[4] For the time being, values are provided only for Polish and English. It is hoped, however, that more languages are going to be covered in the near future with the help of external researchers who are encouraged to submit the collections of values to the authors of the paper. The submission can then be added to the Calculator and published online with a proper reference.

ENGLISH CONSONANTS

OBSTRUENT			SONORANT				VOWEL		
STOP	**AFFRICATE**	**FRICATIVE**	**NASAL**	**LIQUID** 2.0		**GLIDE**			
				lateral	rhotic				
5.0	4.5	4.0	3.0	2.5	2.0	1.0	0		
p b			m			w	1.0	bilabial	**LABIAL**
		f v					1.5	labio-dental	
		θ ð					2.0	inter-dental	**CORONAL**
t d s z			n	l			2.3	alveolar	
ʃ ʒ	ʧ ʤ				ɹ		2.6	post-alveolar	
						j	3.0	palatal	**DORSAL**
k g			ŋ			w	3.5	velar	
							4.0		**RADICAL**
ʔ		h					5.0	glottal	**GLOTTAL**

In the /pr/ example for Polish, MOA for /p/ = 5, MOA for /r/ = 2; POA for /p/ = 1 and POA for /r/ = 2.3. For vowels, MOA is 0.

Then, in order to calculate NAD, the following operation must be performed:

1. /pr/ = |(MOA1 – MOA2| + |(POA1 – POA2)|
2. /pr/ = |5 – 2| + |1 – 2.3| = 3 + 1.3 = 4.3
3. NAD CC = 4.3
4. /rV/ = |MOA1 – MOA2|
5. /rV/ = |2 – 0| = 2
6. NAD CV = 2

/pr/ in Polish is then regarded as preferred as it conforms the condition NAD CC > NAD CV (4.3 > 2).

Analogically, for an English initial cluster /pɹ/, MOA: /p/ = 5, /ɹ/ = 2; POA: /p/ = 1, /ɹ/ = 2.6.

The calculation for the cluster looks as follows:

1. /pɹ/ = |5 – 2| + |1 – 2.6| = 3 + 1.6 = 4.6
2. /ɹV/ = |2 – 0| = 2

/pɹ/ is then regarded as preferred as it also conforms the above condition (4.6 > 2).

The magnitude of the perceptual distance between the two initial consonants in terms of the two criteria corresponds to the preferability of the cluster. Therefore, the /pw/, /bw/

clusters are less preferred than /pj/, /bj/, as the place of articulation of the two member consonants in the former pair is too close to each other (both consonants in the clusters are labial). In contrast, the bilabial and palatal places of articulation are rather distant, which confers a more preferred status upon /pj/ and /bj/.

The NAD conditions for the remaining word positions are specified by the well-formedness conditions listed below. The condition for double final clusters states:

VC_1C_2	NAD (V, C1) ≤ NAD (C1,C2)

Fig. 2. The phonotactic preference condition for double final clusters.

The condition reads:

In word-final double clusters the net auditory distance (NAD) between the two consonants should be greater than or equal to the sonority distance between a vowel and a consonant neighbouring on it.

The condition for double medial clusters states:

$V_1C_1C_2V_2$	NAD (V1, C1) ≥ NAD (C1,C2) ≤ NAD (C2,V2)

Fig. 3. The phonotactic preference condition for double medial clusters.

The condition reads:

For a word-medial double cluster, the NAD between the two consonants should be less than between each of the consonants and its respective neighbouring beat, and it may be equal to the NAD between the first consonant and the beat preceding it.

The predictions for triple clusters in all word positions are formulated below. The condition for triple initials states that:

$C_1C_2C_3V$	NAD (C1,C2) < NAD (C2,C3) ≥ NAD (C3,V)

Fig. 4. The phonotactic preference condition for triple inital clusters.

The condition reads:

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

The condition for triple finals states that:

$VC_1C_2C_3$	NAD (V,C1) ≤ NAD (C1,C2) > NAD (C2, C3)

Fig. 5. The phonotactic preference condition for triple final clusters.

The condition reads:

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

The condition for triple medials states that:

$VC_1C_2C_3V$	NAD (V,C1) ≥ NAD (C1,C2) & NAD (C2, C3) < NAD (C3,V2)

Fig. 6. The phonotactic preference condition for triple medial clusters.

The condition reads:

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

It can be said that the more a cluster diverges from the preference, the more marked (less preferred) it gets.

The values for POA and MOA are the following:

Table 1. Updated MOA and POA values for Polish consonants.

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

Table 2. Updated MOA and POA values for English consonants.

OBSTRUENT			SONORANT				VOWEL		
STOP	AFFRICATE	FRICATIVE	NASAL	LIQUID lateral	LIQUID rhotic	GLIDE			
5.0	4.5	4.0	3.0	2.5	2.0	1.0	0		
p b			m			w	1.0	bilabial	LABIAL
		f v					1.5	labio-dental	
		θ ð					2.0	inter-dental	CORONAL
t d		s z	n	l			2.3	alveolar	
	ʧ ʤ	ʃ ʒ			ɹ		2.6	post-alveolar	
						j	3.0	palatal	DORSAL
k g			ŋ			w	3.5	velar	
							4.0		RADICAL
ʔ		h					5.0		GLOTTAL

Online tool for calculating cluster preferability

Calculating the NAD between two segments involves a series of arithmetic operations. To aid automate calculations several tools have been developed which process input provided by the user and return NAD values along with preferability ratings. The first version of the NAD Phonotactic Calculator (1.0) was developed by Grzegorz Krynicki in the years 2006-2007 (Dziubalska-Kołaczyk and Krynicki 2007). The tool was an online application written in Perl. Version 2.0+, developed by Dawid Pietrala, was largely based on version 1.0+ in terms of functionality and interface design but it was transformed into a desktop application with the option to modify the MOA and POA values for sounds (Pietrala 2014). The most recent version of the tool (3.0), made an online tool again, is available at http://wa.amu.edu.pl/nadcalc/. This version combines the features of the two previous versions: it is an online application which also provides options to modify phone values in the online environment. In addition, it introduces some additional functions such as being able to calculate NAD for several clusters at a time as well as being able to calculate NAD for languages other than English and Polish. It also offers a cleaner and simpler interface.

How to use the online tool

The tool can be used by researchers who work in the framework of B&B to perform calculations of NAD in a robust way. To compute the preferability of a given cluster, the user simply writes it in the input box. The user does not have to take care about the values of POA or MOA as the tool is distributed with a database of default values for specific phonemes as well as with a set of methods which perform the necessary mathematical operations. The application automatically carries out all the calculations and provides the user with the NAD values along with a visual indication of the preferability of the cluster (green or red).

The input can be specified by the user either by writing or copying and pasting the text into the input box or by clicking the IPA symbols located in the table next to the input box. It is possible to calculate the preferability of more than one cluster at a time; this is done by writing one cluster per line. The tool makes use of standard IPA symbols to identify individual consonants in a cluster. In addition, vowels are identified with the use of the V symbol. This is because, at present, the B&B model does not distinguish between vowels when measuring cluster preferability, which is however planned to be covered in the future versions of B&B phonotactics. The inclusion of vowels into the NAD calculation is going to be the next step of the theory extension. Providing the V symbol in a cluster is necessary in order to specify the position of (a) vowel(s) relative to the cluster and, consequently, the word position of the cluster. For example, the string *stV* would be recognised as a word-initial /st/ cluster, while *Vst* would be treated as a word-final /st/ cluster. The tool also recognises word-medial clusters, e.g. *VstV*. Both two- and three-segment clusters are supported in all word positions. The program validates user input so that it returns a user-friendly message when incompatible input is given.

The user can define a number of settings when using the tool. A drop-down list located above the IPA symbols provides a list of supported languages. At the time of writing this paper, the program supports Polish and English as well as a third, user-defined set of parameter values. The default language is English. When the user chooses a language, the program displays the possible segments for that language along with the default POA and MOA values for the segments. The user can make temporary changes to these values if necessary. The modifications of POA and MOA values for Polish and English are saved for the duration of the current session and revert to default when reloading the webpage. The modifications made when using the user-defined option are saved in the web-browser and will be retrieved when accessing the application in the future. It is also possible to control the way NAD is calculated by deselecting the checkbox next to the language selection list. By turning this option off, the program will not take into account sonority when computing the preferability of clusters and will estimate NAD on the basis of MOA and POA.

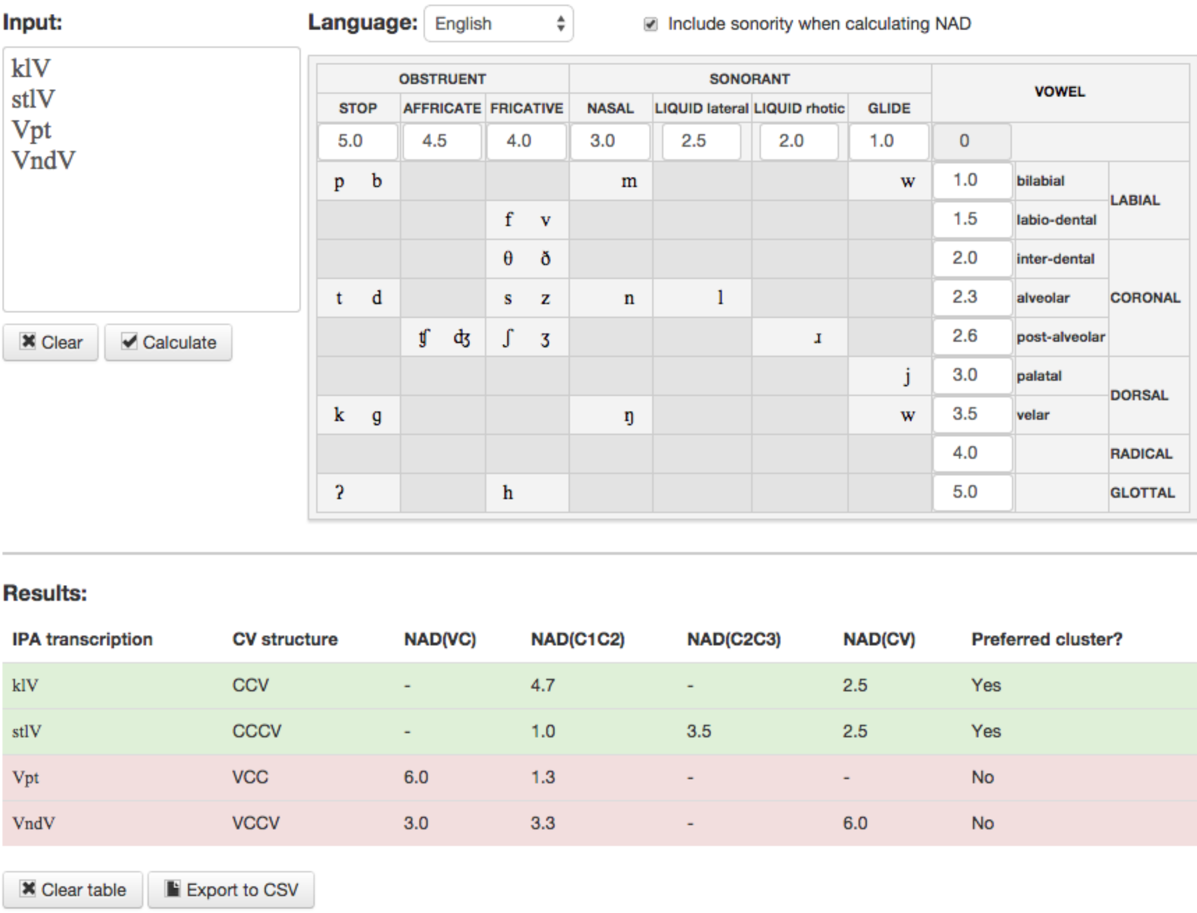


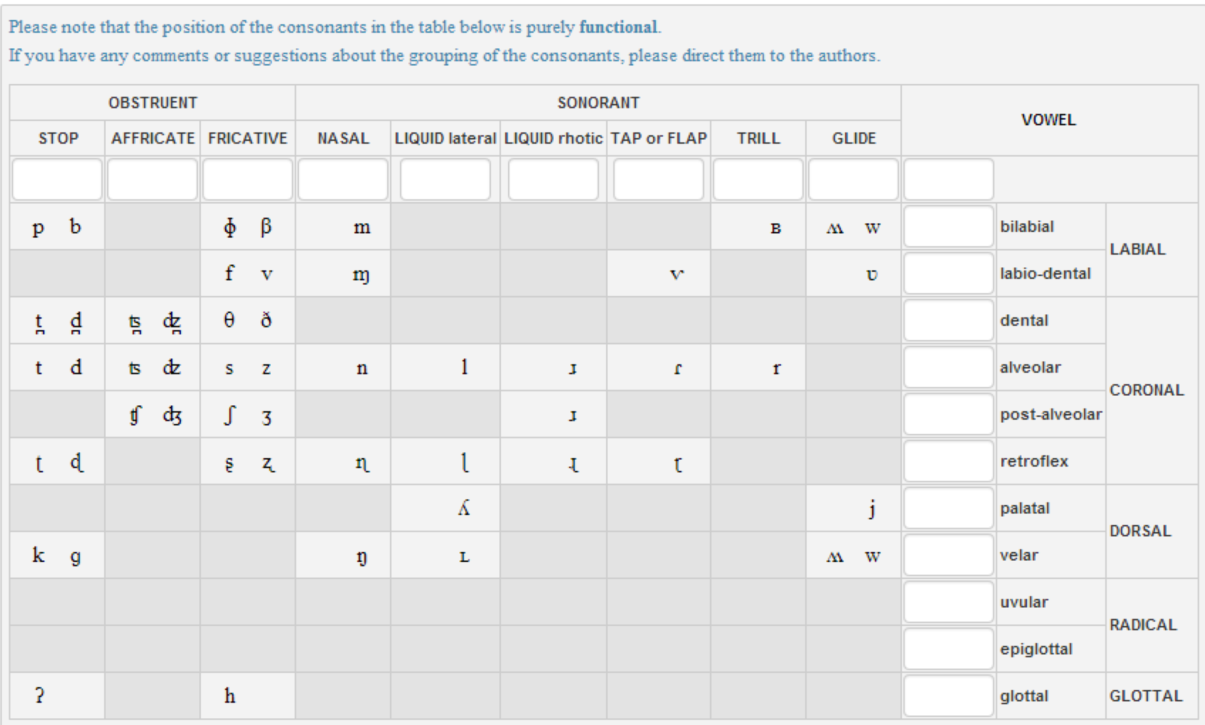
Fig. 7. Phonotactic preference calculator along with sample calculations.

The program operates based on the input provided by the user and the current values of POA and MOA parameters. When the user clicks the *calculate* button underneath the input box, the program will automatically carry out the necessary calculations. First, the program checks whether the segments provided are allowed for the specified language. Then, the CV structure of the cluster is identified based on the number of segments and the position of the vowel(s). Next, the NAD of each pair of segments is calculated on the basis of the currently specified parameter values. Finally, cluster preferability is measured with the help of the corresponding formula. The results are displayed in a table at the bottom of the screen and can be exported to a .csv file.

Considerations to keep in mind

The purpose of the application is to aid researchers in calculating cluster preferability and experimenting with different parameter values for different languages. By offering automatic computation of cluster preferability, the program does away with the laborious task of measuring NAD by hand and limits the possibility of human error in calculations. However, there are some considerations that need to be taken into account when using this tool.

The most important consideration is related to one of the reasons the new version of the programme has been developed: to facilitate computations of cluster preferability for languages other than English and Polish. The user can set their own parameter values for segments present in the language they are researching by choosing the user-defined language option. To keep value manipulations simple, the user is presented with a chart containing the majority of standard IPA symbols. By default, parameter values are empty but can be modified by the user by specifying a number between 0.0 and 5.0 for POA and MOA. Thanks to this, only the values for segments that are relevant for a given language need to be specified. The user-friendliness of this solution, however, comes at a price. First of all, the position of all segments in the chart is fixed and cannot be changed by the user. The positions of segments have been based on places and manners of articulation (Ladefoged 2014, Selkirk 1984). Nonetheless, the order of the segments does not reflect the hierarchy and organisation of segments in all languages. Another limitation imposed by the design is that it is not possible to add new segments that are not covered by the existing chart. Because of the above, the online NAD calculator is not suited to be used with all languages of the world. Users are thus encouraged to contact the authors with comments relating to the layout of the chart as well as with suggested POA and MOA values for new languages.



Please note that the position of the consonants in the table below is purely **functional**.
If you have any comments or suggestions about the grouping of the consonants, please direct them to the authors.

OBSTRUENT			SONORANT						VOWEL		
STOP	AFFRICATE	FRICATIVE	NASAL	LIQUID lateral	LIQUID rhotic	TAP or FLAP	TRILL	GLIDE			
p b		ɸ β	m				ʙ	ʍ w		bilabial	LABIAL
		f v	ɱ			ⱱ		ʋ		labio-dental	
t̪ d̪	t̪s̪ d̪z̪	θ ð								dental	CORONAL
t d	ts dz	s z	n	l	ɹ	ɾ	r			alveolar	
	tʃ dʒ	ʃ ʒ			ɹ					post-alveolar	
ʈ ɖ		ʂ ʐ	ɳ	ɭ	ɻ	ɽ				retroflex	
				ʎ				j		palatal	DORSAL
k g			ŋ	ʟ				ʍ w		velar	
										uvular	RADICAL
										epiglottal	
ʔ		h								glottal	GLOTTAL

Fig. 8. Custom IPA chart with spaces for user-defined parameter values.

Another feature of the software that needs to be considered is its session retrieval capability. All changes made in the user-defined IPA chart are automatically saved in the local storage of the currently used internet browser. Thanks to this solution, user validation is not required to retrieve modifications made to the parameter values. However, these changes will not be carried over when using a different browser or different machine.

One final design choice to mention is the organisation of the output produced by the software. Version 3.0 provides the NAD values for all consonant pairs in the cluster as well as for the first and/or last consonant and the adjacent vowel. Previous version also provided the results of the calculations used to assess cluster preferability. These results could be used to evaluate the relative preferability of two-consontant clusters: the higher the value the more preferable the cluster. However, in the case of three-consonant clusters such evaluations were not possible. To maintain consistency of the results for all cluster types, these calculations are not provided in the output and instead a more general cluster preferability rating is displayed (either 'yes' or 'no') along with colour-coding (green for preferable and red for not preferable). More detailed calculations can be carried out on exported .csv files in software such as Microsoft Excel.

References

Dziubalska-Kołaczyk, Katarzyna. 1995. *Phonology without the syllable. A study in the Natural Framework*. Poznań: Motivex.

Dziubalska-Kołaczyk, Katarzyna. 1997. "Pre- and proto- in Polish phonology and morphology and their interrelations", *Polish Studies in Contrastive Linguistics* 33: 159-171.

Dziubalska-Kołaczyk, Katarzyna. 1999. "Early L1 clusters and how they relate to universal phonotactic constraints", *The Proceedings of the 14th ICPhS. San Francisco*, *1-7 August 1999*, 317-320.

Dziubalska-Kołaczyk, Katarzyna. 2001a. "Phonotactic constraints are preferences", in: Dziubalska-Kołaczyk (ed.), *Constraints and preferences*. Berlin: Mouton de Gruyter, 69-100.

Dziubalska-Kołaczyk, Katarzyna. 2001b. "On the phonotactics of consonantal clusters", in: Chris Schaner-Wolles, John Rennison and Friedrich Neubarth (eds.), *Naturally! Linguistic studies in honour of Wolfgang Ulrich Dressler presented on the occasion of hist 60th birthday*. Torino: Rosenberg & Sellier, 111-119.

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

Dziubalska-Kołaczyk, Katarzyna. 2002b. “Challenges for Natural Linguistics in the twenty first century: a personal view”. *University of Hawai`i Working Papers in Linguistics* 23: 15-39.

Dziubalska-Kołaczyk, Katarzyna. 2002c. “Challenges for Natural Linguistics in the twenty first century: a personal view”, in: Katarzyna Dziubalska-Kołaczyk – Jarosław Weckwerth (eds.), *Future challenges for natural linguistics.* Munich: Lincom, 103-128.

Dziubalska-Kołaczyk, Katarzyna. 2005. “Phonotactics of consonant clusters in the history of English”, in: Antonio Bertacca (ed.), *Historical linguistic studies of spoken English*. Pisa: University of Pisa: PLUS, 15-34.

Dziubalska-Kołaczyk, Katarzyna. 2007. “Natural Phonology: universal principles for the study of language: insiders meet outsiders”, in: Jürgen Trouvain and William J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, 6-10 August 2007*. Saarbruecken: University of Saarbruecken, 71-75.

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. In press. “Are frequent, early and easy clusters also unmarked?”, *International Journal of Linguistics.*

Dziubalska-Kołaczyk, Katarzyna and Grzegorz Krynicki. 2007. Predicting difficulty in second language acquisition. (A paper presented at New Sounds 2007, 25-28 November 2007.) (http://ifa.amu.edu.pl/~dkasia/calculator/) (date of access: 6 Oct. 2008)

Ladefoged, Peter. 2014. *A course in phonetics*. 5th edition. Boston: Heinle&Heinle.

Pietrala, Dawid. 2014. *The Application of Linguistic Principles in Programming: Phonemic Transcription and Phonotactic Constraints*. (Unpublished Ph.D. dissertation). Adam Mickiewicz University in Poznan.

Selkirk, Elisabeth. 1984. On the major class features and the syllable theory, in: Mark Aronoff and Richard T. Oehrle (eds.) *Language sound structure*. Cambridge, Mass.: MIT Press.107-136.