Predicting phonotactic difficulty in second language acquisition

Katarzyna Dziubalska-Kołaczyk[*]
Daria Zielińska

Adam Mickiewicz University, Poznań

dkasia@ifa.amu.edu.pl
daria.zielinska@op.pl

1.   Introduction

The aim of the paper is to demonstrate that universal phonotactic preferences guide the acquisition of consonant clusters in a second language. The empirical evidence comes from young learners of English (L2 English) with mother tongues (L1s) from the following families: independent (Japanese, Korean, Vietnamese), Sino-Tibetan (Chinese), Austronesian (Kosraean, Marshallese, Palauan, Ponapean, Samoan, Tagalog, Trukese, Visayan), Dravidian (Tamil) and Slavic (Polish).

2.   Hypothesis

It is hypothesized that a degree of difficulty in pronouncing L2 clusters would correlate with the universal phonotactic preferences at three levels. At a more general, inventory level, clusters of consonants tend to be avoided, subject to the universal CV preference. The CV is a universal syllable type which occurs in all languages. As shown by Maddieson (2008) on the basis of a sample of 486 languages, 12.5% of the languages allow only CVs, i.e. they have simple syllable structure (C)V. 56.6% of the languages have the moderately complex syllable structure CCVC, with limitations, however, on which consonants may appear in the CC cluster, the second consonant typically being a liquid or a glide. Finally, 30.9% of the languages have complex syllable structure (C)(C)(C)V(C)(C)(C)(C). There is ample evidence showing that more complex structures reduce to CV in various linguistic contexts, e.g., in phonostylistics, first language acquisition, aphasia and other types of language disorders, in language change or in language games and manipulations.

At the word level , medial clusters are tolerated much more than peripheral ones, of which in turn final clusters tolerate longer stretches than initial ones, while at the same time they reduce more heavily (unless inhibited by morphology). Word shape preferences stem both from lexical search/information load criteria and from purely phonetic/phonological criteria (a medial cluster is anchored to vowels on both sides).

Finally, at the level of clusters themselves, less complex clusters are expected to be less difficult than more complex ones, and "better-shaped" ones less difficult than "worse-shaped" ones. Thus, the difficulty for a speaker/learner depends on the characteristics of a given consonantal cluster. In particular, the more universally preferred a cluster, the easier and less susceptible to modifications it is expected to be. NAD (Net Auditory Distance, see section 4) is expected to be a universal criterion, underlying the performance of all subjects,

---

and overriding other relevant factors, such as the structure of the subjects' mother tongue, their experience with English or their other capacities and motivations. The degree of preference for a given cluster is measured by the NAD Principle (see Section 4).

As always in the case of trying to provide evidence for a theory, a failure of any part of the hypothesis will constitute expected feedback to the assumptions underlying the theory. In the case of the present study, it might lead to a potential modification of the form of the phonotactic preferences themselves as well as to an improvement of the NAD model by introducing more refined criteria of measurement.

3. Description of the experiment[i]

53 subjects, aged 11-13 years, native speakers of 15 various languages were recorded, reading 83 times an English carrier sentence *I haven't seen a xxx before!* (e.g., *I haven't seen a kyati before!,... shwepy...,... katewt...,... petewm...,* etc.*)* each time containing a different bi-syllabic nonce word. All words had a trochaic pattern. Each word contained just one double or triple consonant cluster, some of which do appear in English. All word positions (word initial, medial and final) were represented. The clusters consisted of combinations of consonants of all typical sonority values (manners of articulation), so that each sonority value was represented (each of them twice in terms of place of articulation, i.e., similar vs. different) and all consonant tokens were voiceless when available. All initial and final double clusters in the input were acceptable according to the NAD Principle, as well as 70% of the medial doubles. The input triples were fewer and mixed, constituting examples of more and less preferred combinations.

4. Introduction to B&B phonotactics

The present hypothesis stems from a universal model of phonotactics constructed within Beats-and-Binding Phonology (Dziubalska-Kołaczyk, 2002) – a syllable-less theory of phonology embedded in Natural Phonology. The thrust of the theory is the claim that intersegmental cohesion determines syllable structure (Bertinetto et al., 2007), rather than being determined by it (if one insists on the notion of the "syllable" which is epiphenomenal here).

Phonotactic preferences specify the universally required distances between segments within clusters which guarantee, if respected, preservation of clusters (cf. intersegmental cohesion). Clusters, in order to survive, must be sustained by some force counteracting the overwhelming tendency to reduce towards CVs (CV preference). This force is a perceptual contrast defined as the Net Auditory Distance Principle (NAD Principle) (cf. Dziubalska-Kołaczyk, 2002, 2003; Dressler & Dziubalska-Kołaczyk, 2006; Dziubalska-Kołaczyk & Krynicki, 2007a).

$$NAD = |MOA| + |POA| + |Lx|$$

whereby MOA, POA and Lx are the absolute values of differences in the Manner of Articulation, Place of Articulation and Voicing of the neighbouring sounds respectively.

Let us consider an example of the preference concerning word-initial two-consonant clusters:

$$NAD\ (C1,C2) \geq NAD\ (C2,V)$$

The preference 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 distances in terms of manner and place of articulation are calculated on the basis of the table below (Table 1). The manners and places assumed in the table are selected

according to their potential relevance: 6 manners (stop, affricate, fricative, sonorant stop, approximant, semivowel) where affricates and semivowels are attributed half a distance due to their dubious 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, p. 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. In fact, one would need to investigate from the auditory perspective as many acoustic/articulatory cues as possible which potentially contribute to the overall perceptual impression brought about by phonotactic sequences. This, however, is a wider research perspective reserved for future investigation. In the present research and for the purposes of the present data, the assumption has been made as described above and in Table 1.

Table 1. Distances in manner and place of articulation.

4	3	2	1	0	
obstruent		Sonorant			
stop	fricative	sonorant stop	approximant	V	
	affricate			semiV	
				labial	1
				coronal	2
				dorsal	3
				radical	4
				laryngeal (glottal)	5

Consider again the preference for initial double clusters:
NAD (C1,C2) ≥ NAD (C2,V)
Let us now define two Net Auditory Distances between the sounds (C1, C2) and (C2, V) where
- C1   (MOA1, POA1, Lx1)
- C2 (MOA2, POA2, Lx2)
- V   (MOA3, Lx3)

in terms of the following metric for (C1, C2) cluster

|MOA1 - MOA2| + |POA1 - POA2| + |Lx1 - Lx2|

and

|MOA2 – MOA3| + |Lx2 – Lx3|

for (C2, V) cluster.

Let us now exemplify the above with an English initial double cluster in the word *try*.

$$t = (4, 2, 0), r = (1, 2, 1), V = (0, 0, 1)$$
$$\text{NAD (C1, C2)} = |4\text{-}1| + |2\text{-}2| + |0\text{-}1| = 3+0+1=\textbf{4}$$
$$\text{NAD (C2, V)} = |1\text{-}0| + |1\text{-}1| = 1+0=\textbf{1}$$

Thus, the preference NAD (C1,C2) ≥ NAD (C2,V) is observed because **4** > **1**.

The NAD Principle makes finer predictions than the ones based exclusively on sonority; for instance, it shows that among stop+liquid initial clusters, prV and krV > trV, brV, grV > drV, etc. (since their NAD's are respectively: 5 > 4 > 3). This universal principle leads to predictions about language-specific phonotactics, its acquisition and change. Specifically, it also makes it possible to predict and explain the order of difficulty in the acquisition of second language phonotactics which appears to be universally valid and as such calls for similar remedies across languages. For example, if one compares the frequent English (Fig. 1) and Polish (Fig. 2) clusters, one can observe that among the English ones many more clusters are universally preferred (i.e., they observe the respective preference for initial doubles discussed above). A Polish learner of English is therefore expected to have fewer difficulties in the acquisition of those clusters than an English learner of Polish.
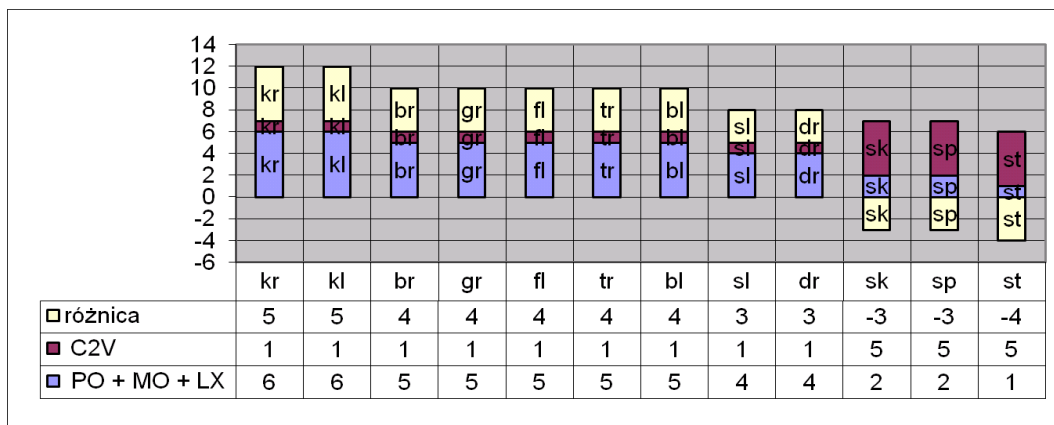

	kr	kl	br	gr	fl	tr	bl	sl	dr	sk	sp	st
□ różnica	5	5	4	4	4	4	4	3	3	-3	-3	-4
■ C2V	1	1	1	1	1	1	1	1	1	5	5	5
■ PO + MO + LX	6	6	5	5	5	5	5	4	4	2	2	1

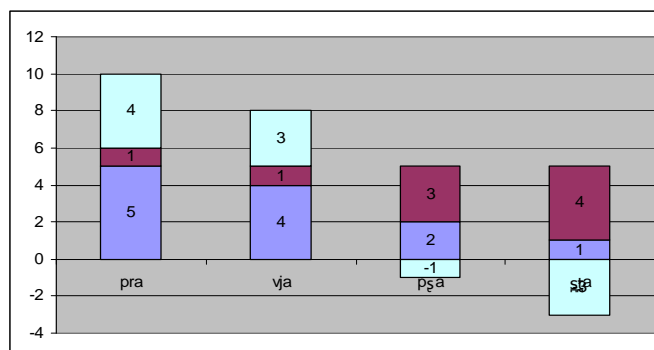**Figure 1**. Frequent English initial doubles according to the NAD Principle.

Figure 2. Selected frequent Polish clusters according to the NAD Principle.

5. Phonotactic calculator[1]

For the purposes of B&B phonotactics, a prototype of a phonotactic calculator has been developed (Dziubalska-Kołaczyk & Krynicki, 2007b). Its purpose is to make it possible to perform a statistical analysis of large quantities of phonetic data, such as phonetic dictionaries and phonetically annotated corpora from various languages. The calculator will work on various lengths of clusters in all word positions (initial, medial and final) and estimate them with respect to the universal phonotactic preferences/constraints formulated to define each cluster type in each position (cf. Dziubalska-Kołaczyk, 2002, Section 4 above). It provides fast feedback on the predictability value of those preferences/constraints.
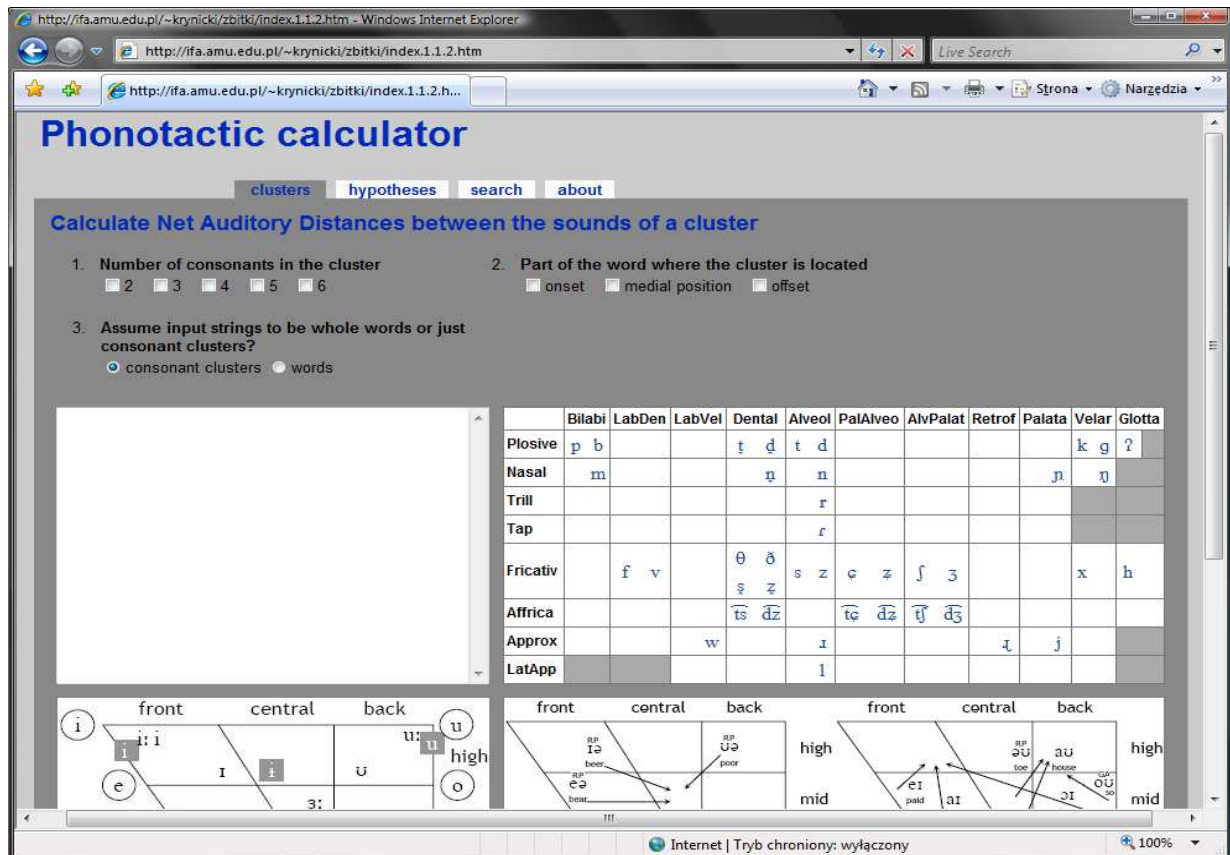
Figure 3. Phonotactic calculator: view of the website.
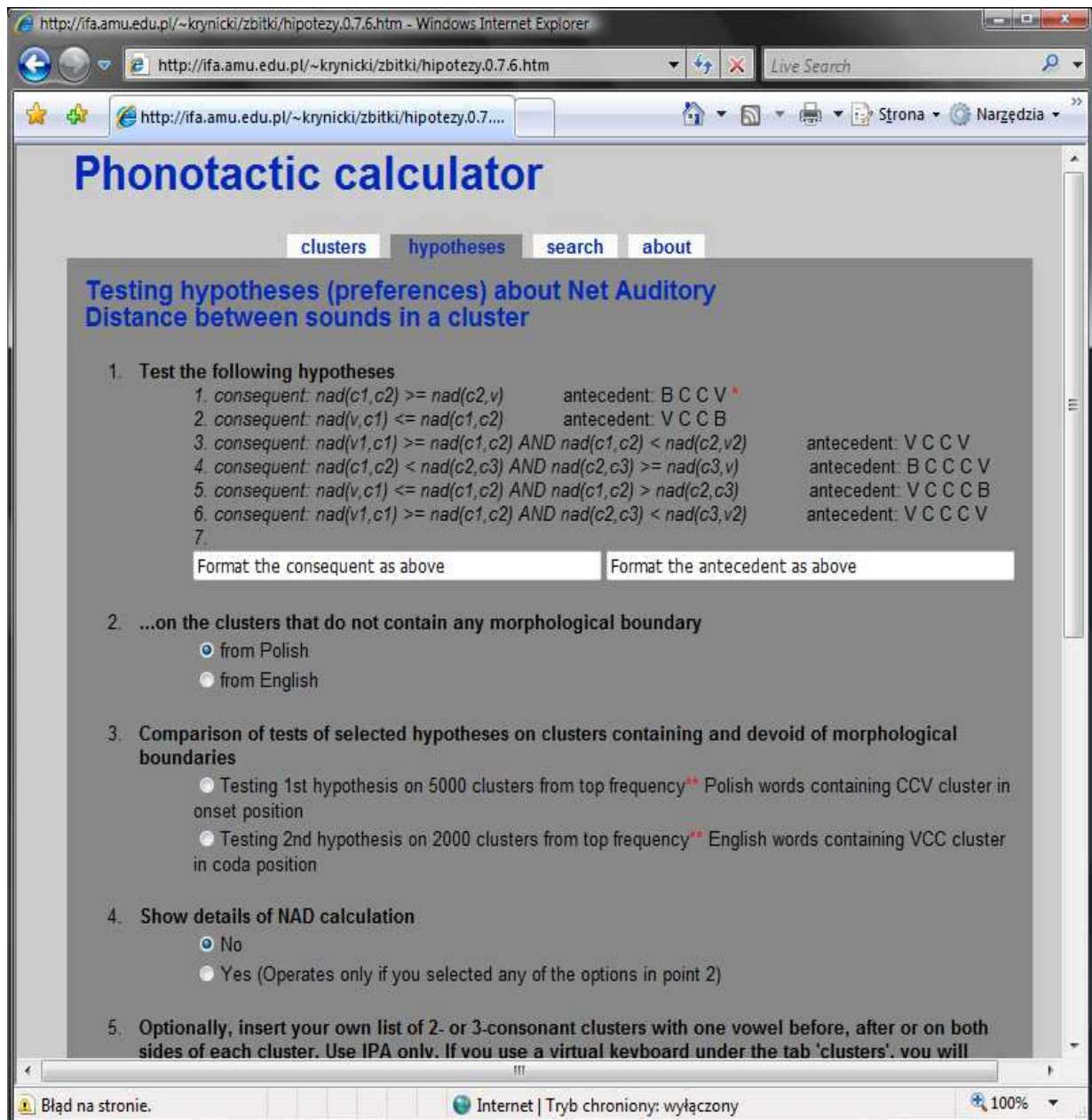
Figure 4. Phonotactic calculator: view of the website cont.

6. Analysis of the data

A total of 4399 utterances were analyzed, produced by 53 children, each reading 83 sentences containing a nonsense word with a 2- or 3-consonant cluster. In 3031 of these utterances (68.9%) the speakers modified or avoided the cluster that was assumed to be the correct pronunciation of the nonsense word. Both double and triple clusters produced by the speakers can be analyzed according to many different criteria. In the article, however, we will focus on the selected data which comes solely from the renderings of the two-consonant clusters. Nevertheless, it is worth mentioning that 85 (10.69%) out of 795 utterances produced containing a triple cluster were pronounced correctly (i.e., exactly as the expected consonant cluster), whereas in 593 (74.59%) of those renderings there were too many changes both in the phoneme structure and the length of the word to make their systematic classification possible.

6

The procedure

The subjects' utterances were analyzed auditorily by the second author, with some prior analysis by Dziubalska-Kołaczyk and Krynicki (2007c) available for comparison. The following error types were distinguished and transcribed (Table 2):

Table 2. Types of errors classified.

error description
substitution of one or two consonants in the cluster by a consonant or consonants not present in the expected cluster
reordering of consonants in the cluster
pause insertion between the elements of a consonant cluster
change of the expected double cluster into a CV(C) sequence
change of the expected double cluster into a triple cluster by adding a new phoneme to the cluster
cluster status change (e.g. from initial to medial)
unintelligible pronunciation
substantial mispronunciation of the word
omission of the word

General results for double consonant clusters

There were 68 nonsense words with a double consonant cluster used in the experiment. Table 3 shows the number of clusters occupying a given position in the word and the number of clusters following phonotactic preferences according to the NAD Principle in the input.

Table 3: Cluster position and phonotactic preference.

preference	number of cluster that apply to a given preference	number of clusters that follow the phonotactic preference	percentage
initial	17	17	100%
final	13	13	100%
medial	38	27	71%

The subjects produced 3604 word tokens that were supposed to contain a double consonant cluster. 64% of them (2321) were different from the expected cluster (see Fig. 5).
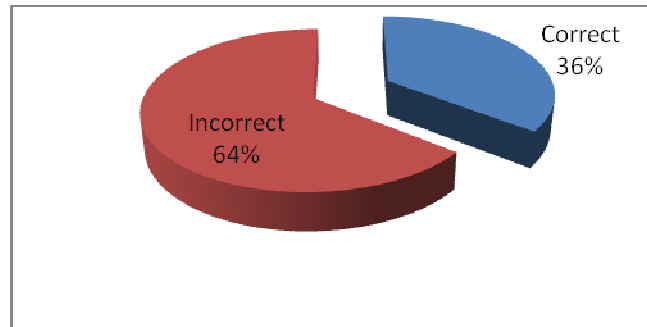
Figure 5: Doubles: realizations.

Figure 6 presents different strategies to which the speakers resorted in order to avoid producing the expected cluster.
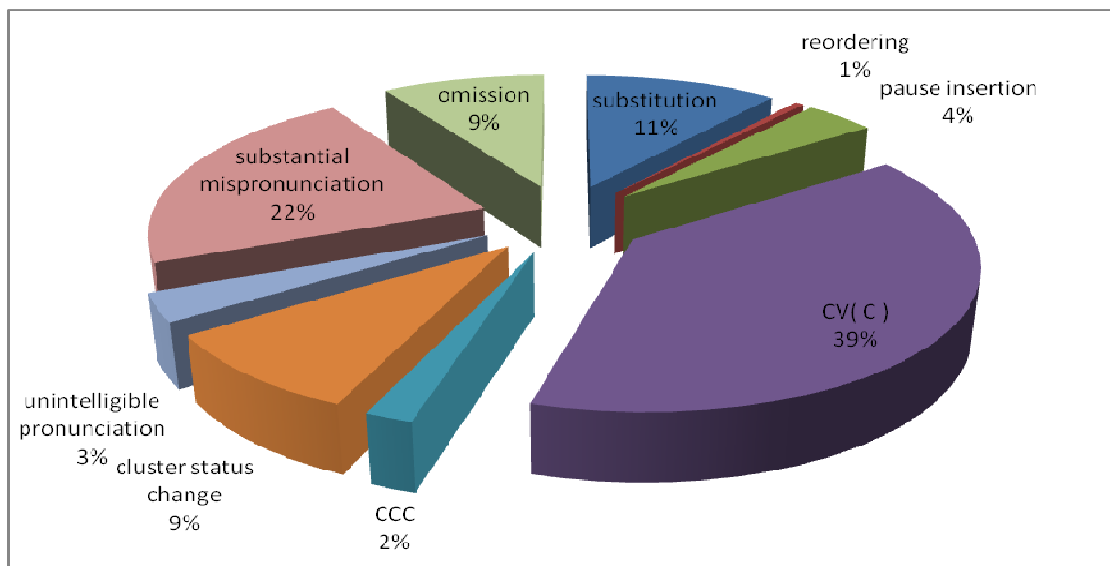

Figure 6: Doubles: types of errors.

In 212 of the utterances (9.14%) the status of the cluster was changed (e.g., from medial to final). Figure 7 shows that 85% of all new clusters created in the process of cluster status change were medial ones, whereas Figure 8 demonstrates that although 98% of the new medials do not follow phonotactic preferences according to the NAD Principle, it is still preferable to have even a phonotactically dispreferred medial cluster than a phonotactically preferred peripheral one.
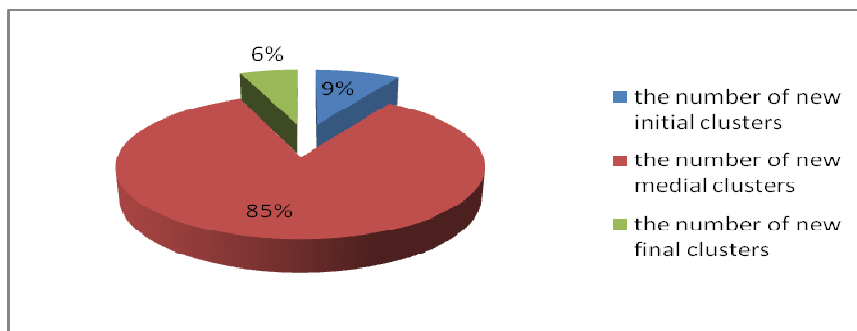
Figure 7: Cluster status change

8
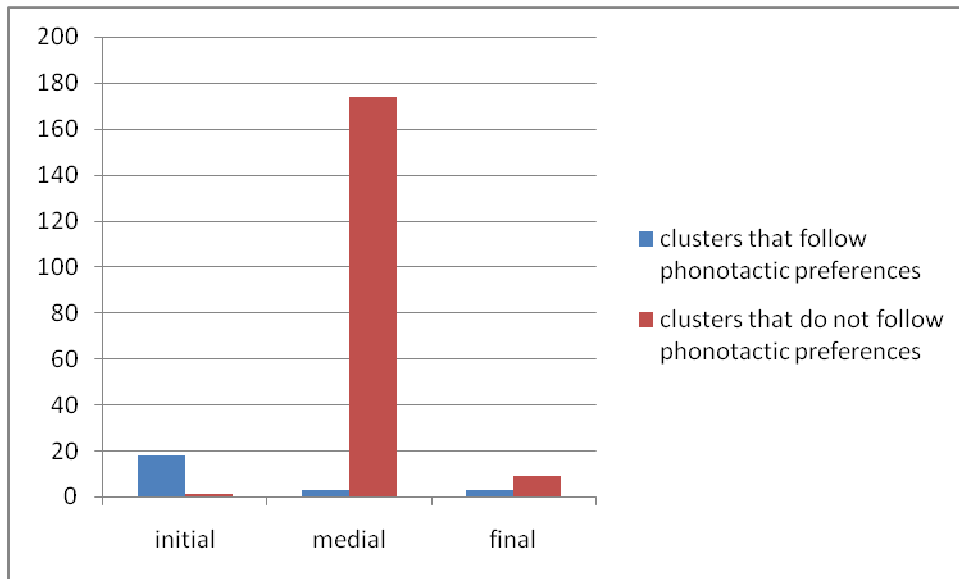
Figure 8: Phonotactic preferences of the new clusters.

Initial double clusters

The speakers produced 901 utterances that were supposed to contain a double initial cluster. 628 of the utterances (70%) were modified by the subjects (see Fig. 9).
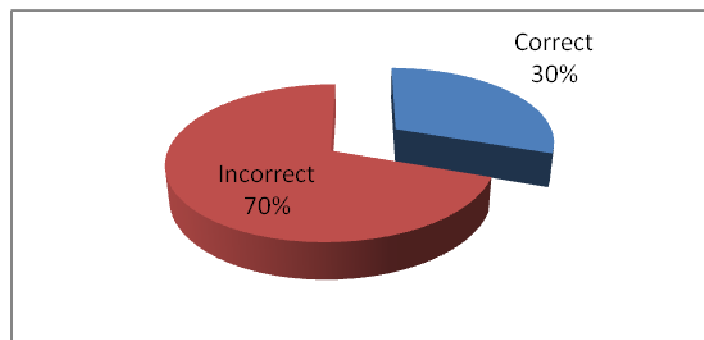

Figure 9: Initial doubles: realizations.

Figure 10 presents the types of errors that were made in the incorrect renderings of 2-consonant initial clusters.
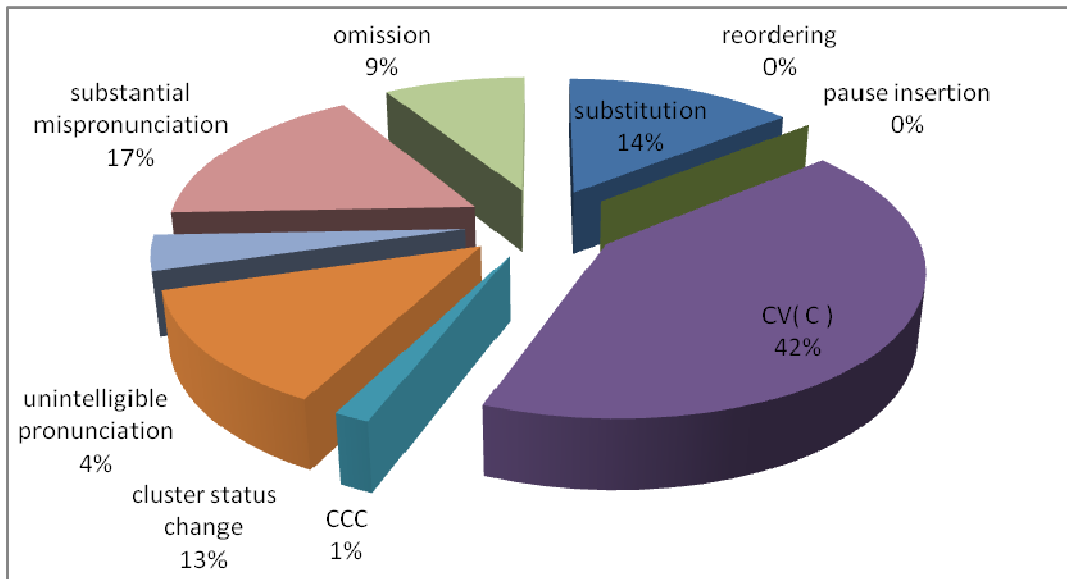
Figure 10: Initial doubles: types of errors.

Figure 11 is a demonstration of the strategies by means of which the speakers broke the cluster into a CVC sequence. Importantly, in 55.73% of the utterances the speakers preserved the input, hence vowel insertion significantly outnumbers reduction.
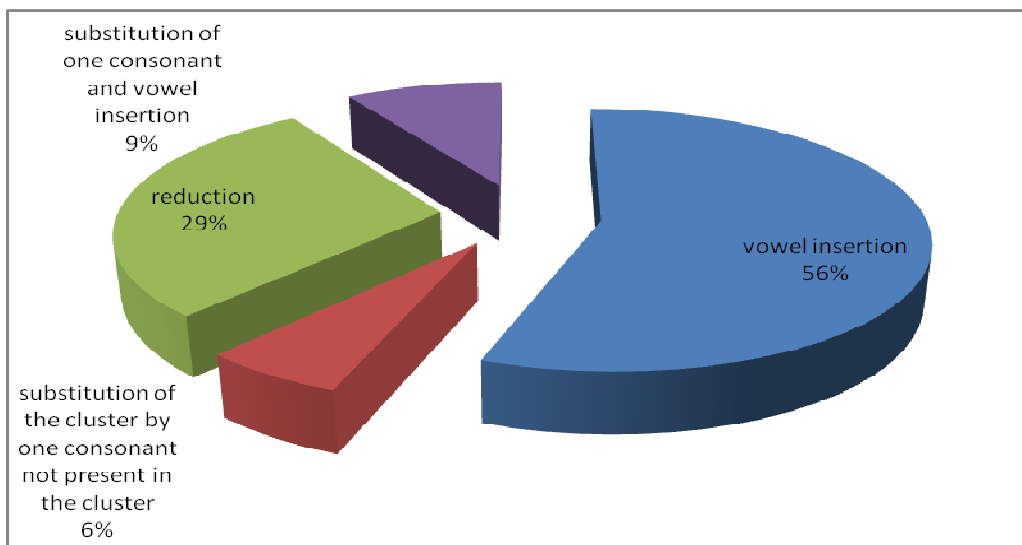

Figure 11: Cluster modification towards CVC.

The subjects were supposed to read 17 nonsense words with double initial clusters, all of which follow phonotactic preferences according to the NAD Principle. 362 double initial clusters were elicited, 96% of which still followed the preference.

The data shows that there is a correlation between NAD (CC) – NAD (CV) and the number of incorrect renderings of the initial double clusters (see Fig. 12).
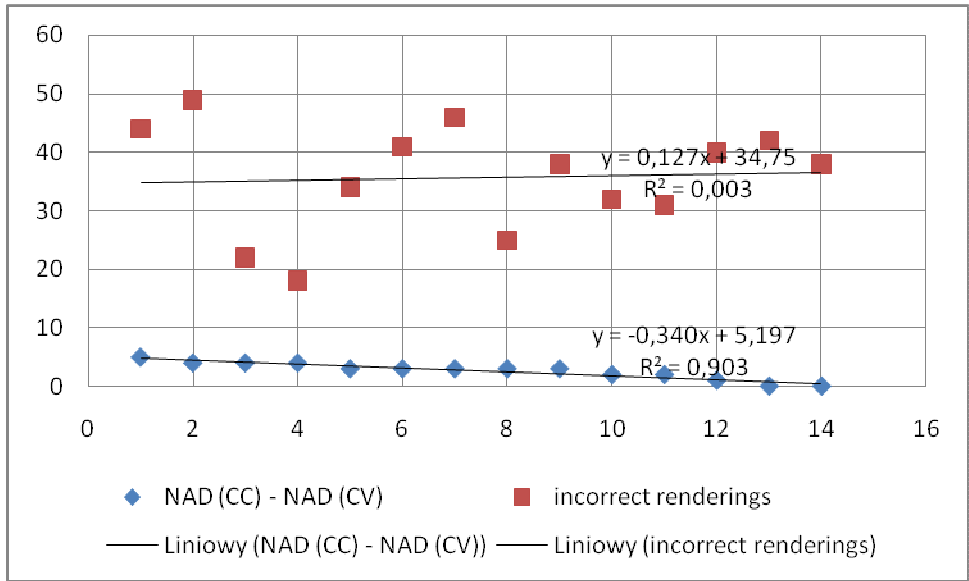
Figure 12: Correlation between NAD (CC) – NAD (CV)
and incorrect renderings.

Figures 13, 14 and 15 demonstrate a strong correlation between NAD (CC) – NAD (CV) and the number of cluster status changes, cluster modifications towards CVC and reduction errors respectively.
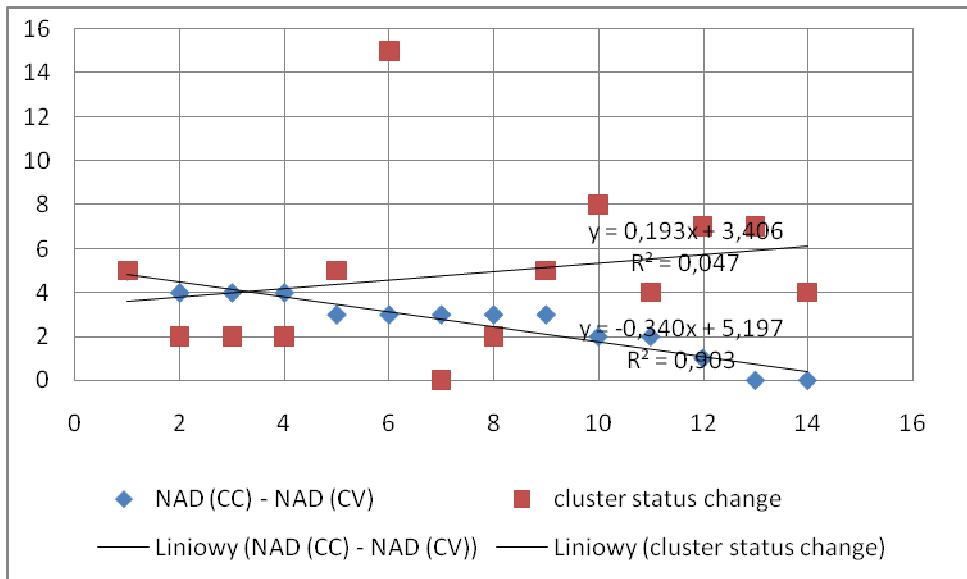

Figure 13: Correlation between NAD (CC) – NAD (CV)
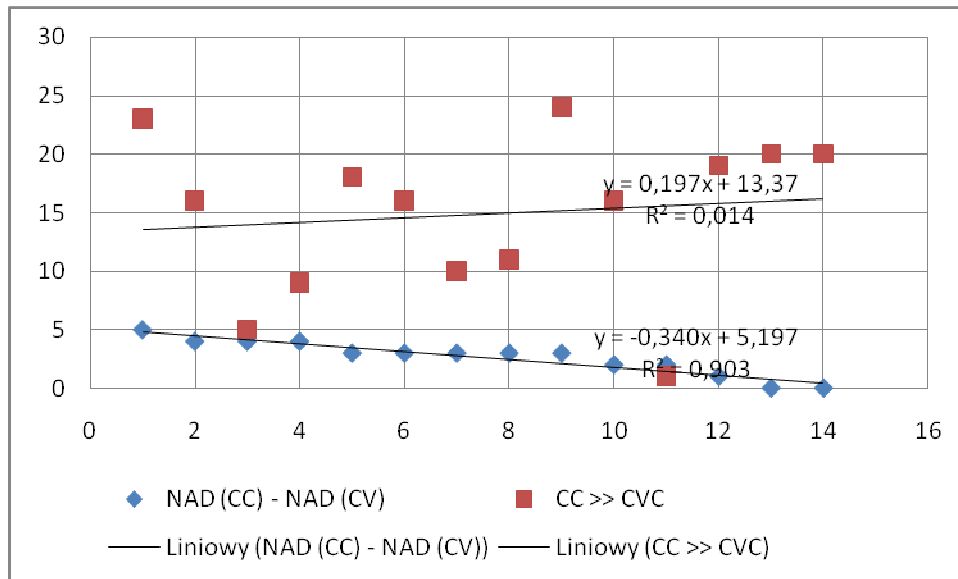and cluster status changes.

Figure 14: Correlation between NAD (CC) – NAD (CV)
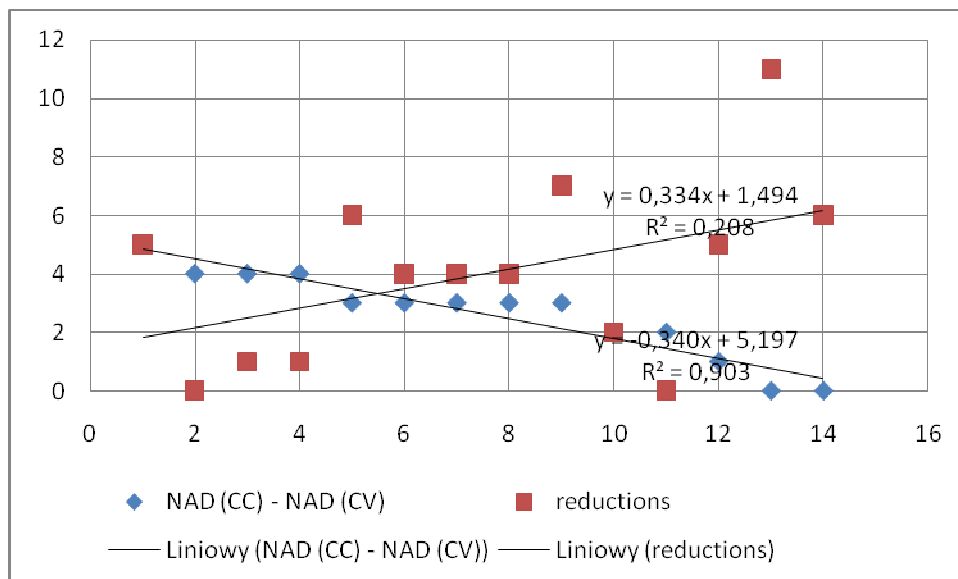and cluster modifications towards CVC.

Figure 15: Correlation between NAD (CC) – NAD (CV)
and reduction errors.

Medial double clusters

A total of 2014 renditions that were supposed to contain a medial double cluster were obtained in the experiment. 1167 of the utterances (58%) diverged from the expected ones (see Fig. 16), which is significantly less than in the case of both initial and final clusters.
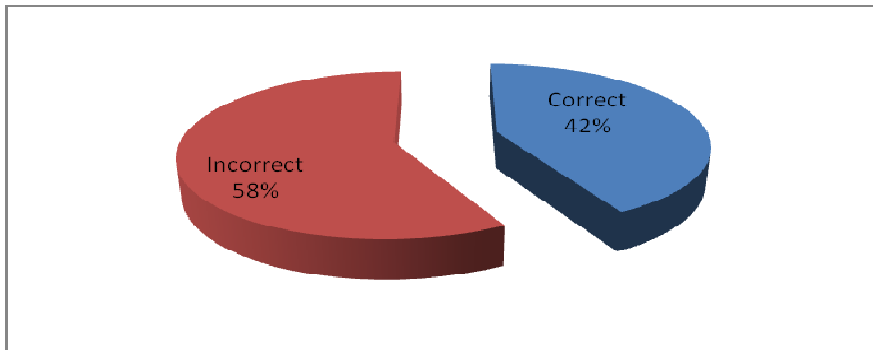
Figure 16: Medial doubles: realizations.

Figure 17 is a graphic representation of the types of errors in the renderings of medial double clusters.
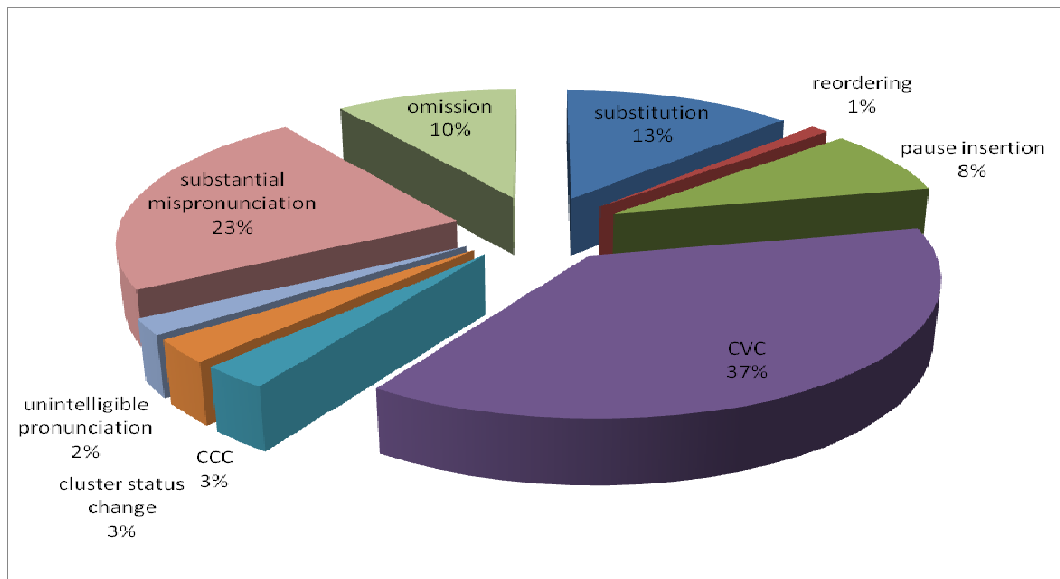

Figure 17: Double medials: types of errors.

Out of 38 nonsense words with a double medial cluster, 27 were considered phonotactically preferred. The subjects rendered 669 word tokens with a double medial that was expected to follow the preference. Again, the majority of them (97%) were "good" from the point of view of the NAD Principle.

Figure 18 shows the realizations of medial double clusters that according to NAD are phonotactically preferred. It is noteworthy that geminates have relatively few correct renditions in comparison to the clusters whose components have different places of articulation.
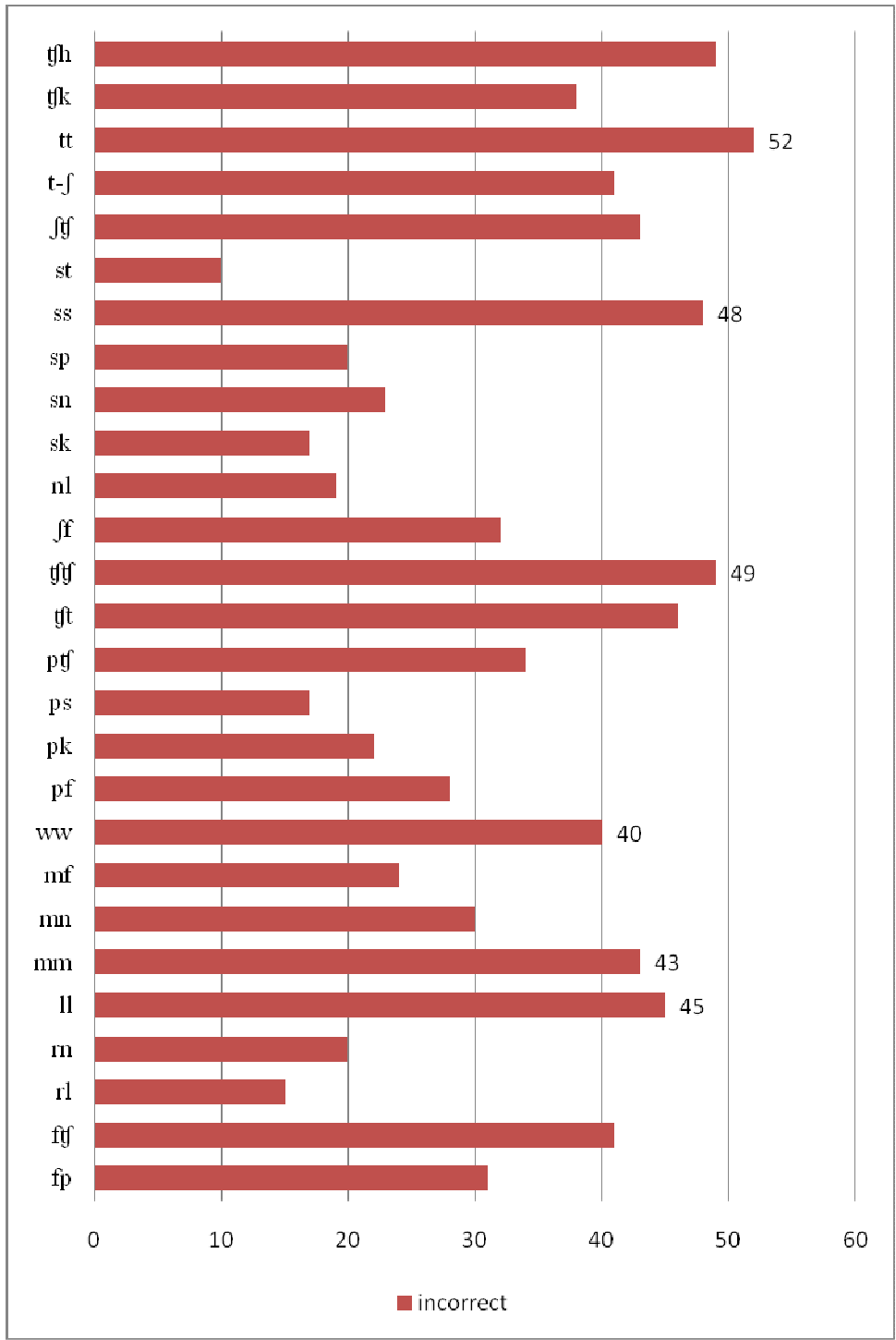
Figure 18: preferred medials: realizations

Final double clusters

The speakers produced 689 utterances that were supposed to contain a double final cluster. 526 of the utterances (76%) were incorrect (see Fig. 19).
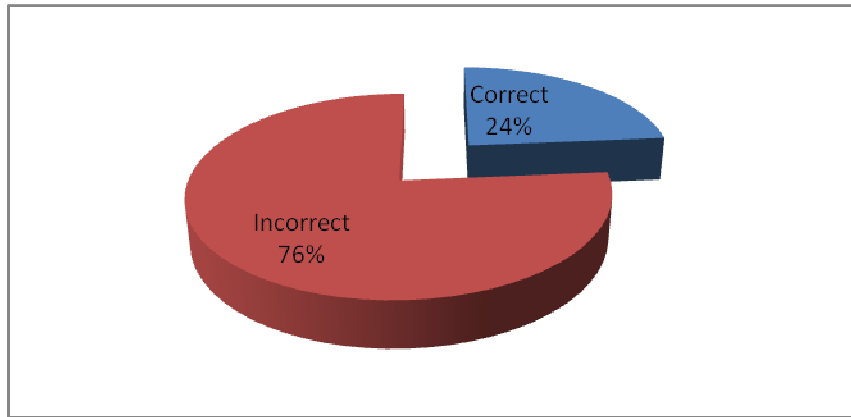
Figure 19: Double finals: realizations.

Figure 20 is a presentation of errors made by the subjects in the renditions of double final clusters.
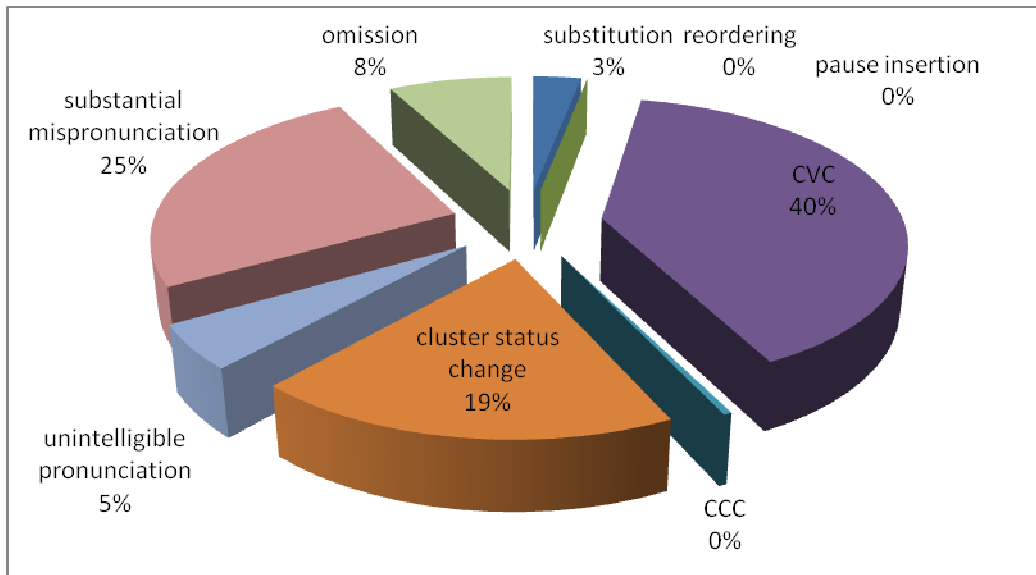

Figure 20: Final doubles: types of errors.

Figure 21 demonstrates the errors that the subjects made while modifying the cluster into a CVC sequence. Unlike in the case of the initial clusters, there were considerably more reductions than vowel insertions.
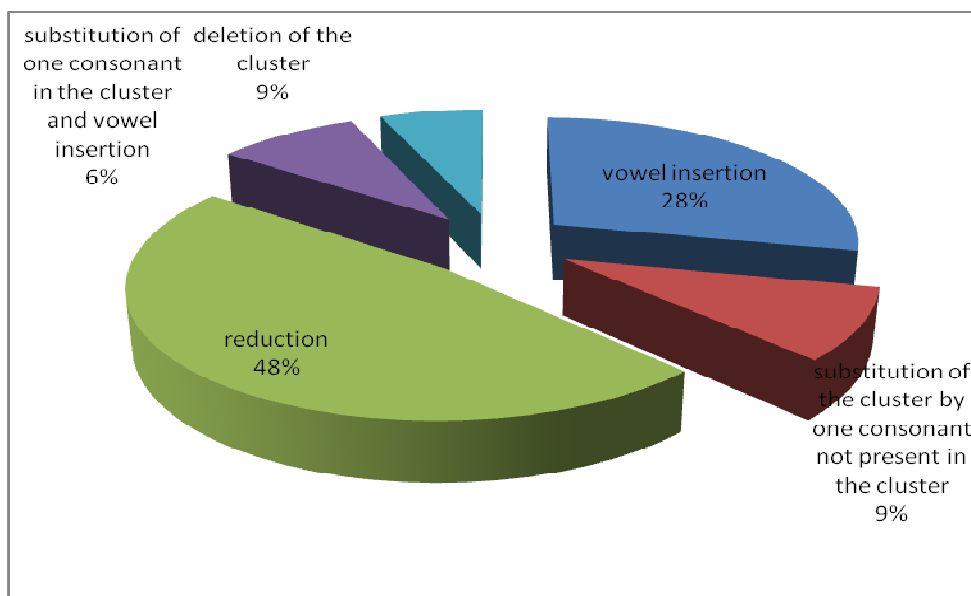
Figure 21: Cluster modification towards CVC.

The speakers were supposed to read 13 nonsense words with double final clusters, all of which follow phonotactic preferences according to the NAD Principle. 179 double final clusters were rendered, 99% of which still followed the preference.

Figure 22 presents a correlation between NAD (VC) – NAD (CC) and the number of incorrect renderings of the final double clusters.
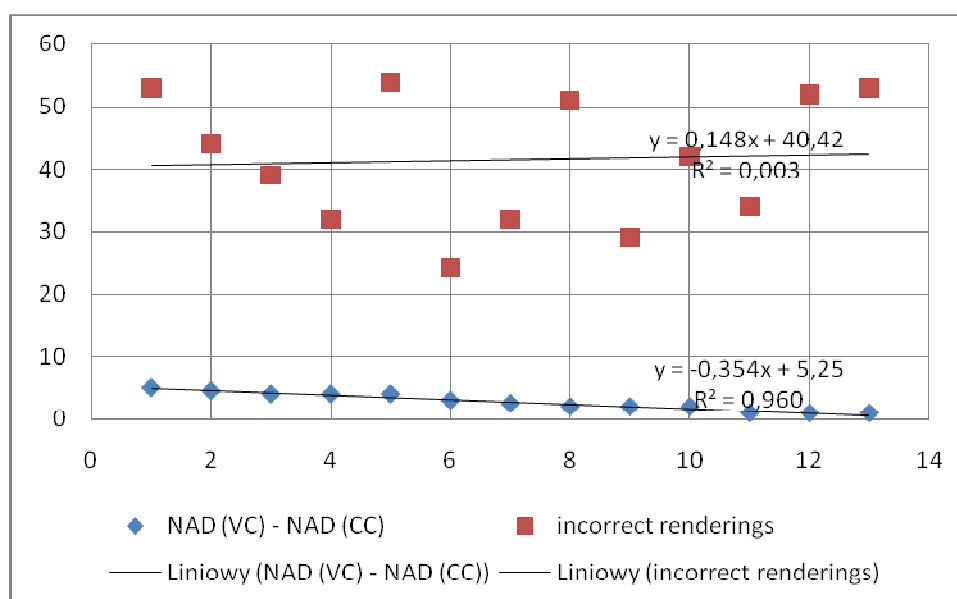

Figure 22: Correlation between NAD (VC) – NAD (CC)
and incorrect renderings.

NAD (VC) – NAD (CC) turns out to have a statistically significant influence on the number of final double clusters reduced to CVC (see Fig. 23).
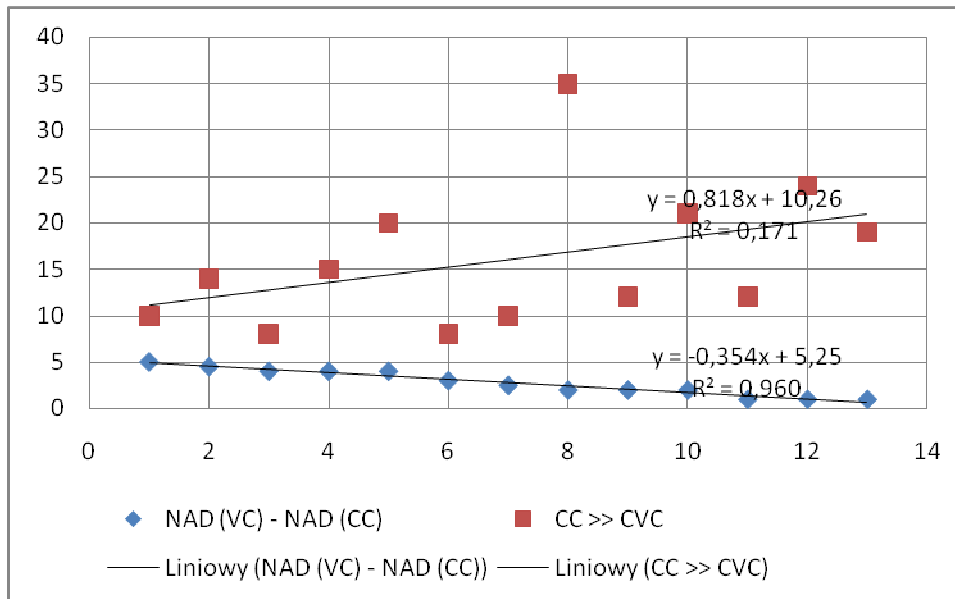
Figure 23: Correlation between NAD (VC) – NAD (CC) and
modifications towards CVC.

The analysis of the data also showed that there is a correlation between NAD (VC) – NAD (CC) and the number of reduction errors that the subjects made (see Fig. 24).
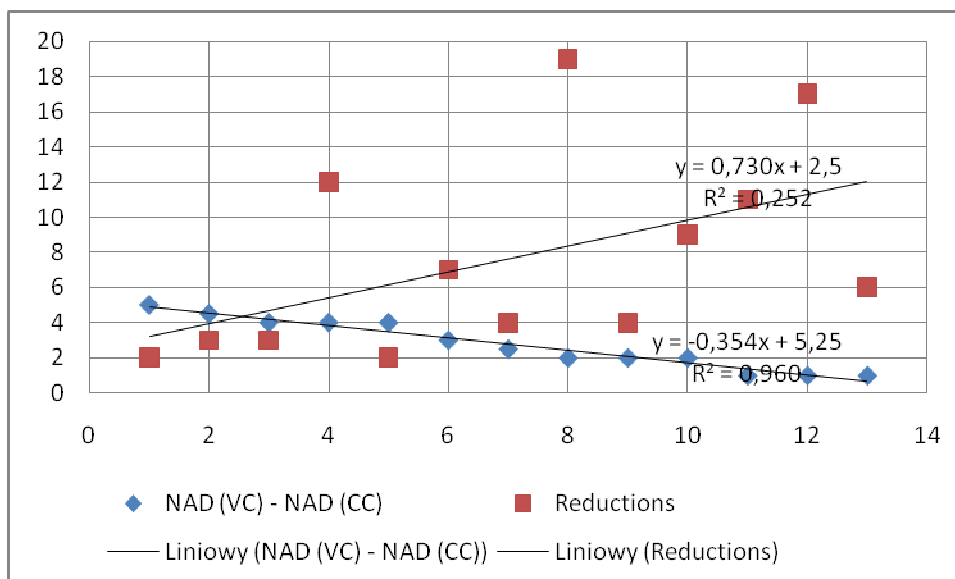
Figure 24: Correlation between NAD (VC) – NAD (CC) and reductions.

7. Summary and conclusions

Let us interpret the obtained results with reference to the three-level hypothesis formulated in section 2. At the most general level, it was predicted that clusters as such would be avoided or at least struggled with. Indeed, 68.9% of all 2-consonant input clusters were not produced correctly by the subjects. They were modified in various ways described above, in order to make their pronunciation easier. Only 22.54% of the modifications resulted in producing an altered cluster, yet a cluster after all. 3-consonant clusters, predictably, scored still worse, since 74.59% of the attempted productions failed, and the failure was much more

profound than in the case of double clusters. Thus, at the most general level the hypothesis holds.

At word level, initial, medial and final clusters were predicted to behave in accordance with their different status. Only 58% of the medial clusters were rendered incorrectly, which is much more successful when compared to 70% incorrect initial clusters and 76% incorrect final ones. The higher tolerance for medial clusters was confirmed by the changes of the other two "positions" of a cluster into a medial position – 85% of the cases.

In order to break a cluster into a CVC, the subjects used vowel epenthesis in 56% of the cases in initial clusters, while only in 28% of the cases in final clusters. Reduction of one of the consonants in order to obtain a CV(C) was used in reverse proportion: 29% of the times in initial clusters and 48% of the times in final ones. This clearly points to the tendency to preserve more of the input in the initial than in the final position of the word. This tendency stems from the semiotic salience of the word-initial position, facilitating lexical search and carrying high informational load. Word-final position is more susceptible to reduction due to its lesser salience unless, of course, it acquires a morphological function (which, however, was not the case in the test words). This is further confirmed by more substitutions (13%) initially (to preserve a cluster) than finally (3%), and fewer changes of an initial into a medial cluster (13%) than of a final into a medial one (19%). Summing up, at word level the hypothesis is also confirmed.

Finally, let us concentrate on the phonotactic level itself. Certainly it has been confirmed that less complex clusters are easier for learners than more complex ones. 36% of 2-consonant clusters vs. only a little over 10% of 3-consonant clusters were correctly produced by the subjects. The longer the cluster the more difficult it gets. Additionally, some type of repair was significantly much more successful in rendering the shorter clusters.

The most detailed and at the same time the strongest claim concerned the predictability power of the NAD Principle. The "better-shaped" clusters were expected to be proportionally easier for the learner than the gradually "worse-shaped" ones. In other words, the difficulty of a cluster was expected to correlate with the "goodness" of a cluster, i.e., the degree of universal preferability of a cluster. The results obtained confirm the above claim in a number of interesting ways. Firstly, and more generally still, a medial cluster is always more acceptable, since a NAD-dispreferred medial is easier than NAD-preferred initials or finals (Fig. 8). Secondly, a definite majority (96%, 97% and 99% respectively) of the modified clusters <u>preserved</u> their NAD-preferred status; i.e., although the input cluster originally posed an (unexpected) difficulty, it was still modified so that it continued to observe the universal preference. Thirdly, correlations have been found for both initial and final clusters between their degree of preferability ("goodness") and the proportional degree of difficulty in their production (the number of errors). In particular, the less preferred a cluster the more difficult it appeared (the strongest correlation showing for the degree of reduction of a difficult cluster). Therefore, NAD proves to be quite a reliable measure for universal phonotactics.

The expected constructive feedback to the theory was predominantly positive, as discussed above. Additionally, it was noticed that the NAD measurement is able to make finer predictions than sonority-only based phonotactics (the correlations would have been less prominent or nonexistent). We cannot develop the issue further here for lack of space. Quite prominent negative feedback emerged from the analysis of medial clusters. In particular, geminates turned out not to be the most preferred medials; rather, combinations of consonants differing by one place or manner feature prevailed (significantly, <st> was the most preferred). This result calls for a modification of a universal preference for medials: rather than claiming the smallest C1C2 distance possible, one would need to introduce a necessary minimal distance there. It remains to be investigated how to define it.

References

Bertinetto, P. M.; Scheuer, S.; Dziubalska-Kołaczyk, K., & Agonii, M. (2007). Intersegmental cohesion and syllable division in Polish. Extended version. In *Reports of Laboratorio di Linguistica*, Scuola Normale Superiore, Pisa. http://linguistica.sns.it

Dressler, W. U., & Dziubalska-Kołaczyk, K. (2006). Proposing morphonotactics. *Rivista di Linguistica*, *18* (2), 249-266.

Dziubalska-Kołaczyk, K. (2002). *Beats-and-Binding Phonology*. Frankfurt am Main: Peter Lang.

Dziubalska-Kołaczyk, K. (2003). On phonotactic difficulty. *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2729-2732). Barcelona.

Dziubalska-Kołaczyk, K., & Krynicki, G. (2007a). Phonotactic preferences in Polish, English and German: quantitative perspective. *Poznań Linguistic Meeting 2007*. Conference presentation.

Dziubalska-Kołaczyk, K., & Krynicki, G. (2007b). http://ifa.amu.edu.pl/~dkasia/calculator/

Dziubalska-Kołaczyk, K., & Krynicki, G. (2007c). Predicting phonotactic difficulty in second language acquisition. *Talk at New Sounds 2007*. Florianópolis.

Ladefoged, P. (2006). *A Course in Phonetics*. 5th edition. Boston: Heinle & Heinle.

Maddieson, I. (2008). Syllable structure. In M. Haspelmath, & M. S. Dryer; D. Gil, & B. Comrie (Eds.), *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library, Chapter 12. Available online at http://wals.info/feature/39. Retrieved on 2 Jan 2009.

---

[i] This empirical project constitutes part of a larger research programme which the first author conducted at the University of Hawai'i at Mānoa as a Senior Fulbright scholar in the academic year 2001-2002. The subjects were ESL students from Washington Middle School (grades 6-8) in Honolulu. The recordings took place on the school premises thanks to the courtesy of the headmistress of the school, to whom I would like to express my gratitude and thanks. Upon my application, the project received the Exempt Research status from The Committee for the Protection of Human Subjects (CHS), which is the unit designated to function as the federally mandated Institutional Review Board (IRB) for the University of Hawai'i (UH) system. CHS is responsible for the compliance review and the overseeing of all research protocols affiliated with the University of Hawai'i that involve the use of human subjects.