

Influence of Zipfian distribution on learning second order phonotactic constraints

Laura Baakman

May 10, 2013

Abstract

Previous work (Onishi et al. 2002) has shown that adult subjects can learn novel second order phonotactic constraints in an artificial language. This paper investigates if second order phonotactic constraints in an artificial language are better learned when trained with a Zipfian distribution than with a uniform distribution. Subjects listened to CVC words with second order constraints. To test their knowledge of the learned words they then listened to and repeated a superset of the learned words. This revealed no benefit of the Zipfian distribution over the uniform distribution. No evidence of second order constraint learning was found. Possible explanations are that we used too few presentations or that our language was too complex.

1 Introduction

In earlier research it has been shown that adults can learn the phonotactic regularities of syllables for an unknown language after only a brief auditory presentation when the words are presented in a uniform distribution. The research described in this paper investigates if adults acquire these regularities better when certain tokens are presented more often than others, e.g. with a Zipfian distribution.

Words consist of syllables and, syllables consist of at most three constituents. All syllables contain a nucleus, which is the vowel-like element. The nucleus may be preceded or followed by a string of consonants, the onset and the coda respectively (Fromkin et al. 2000). A syllable can consist of only a nucleus. e.g. the English word ‘I’. A two-syllable word, such

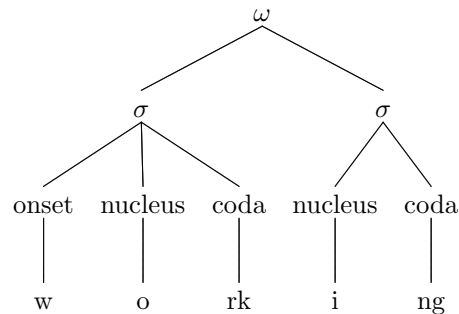


Figure 1: The diagram of the English word ‘working’, ω denotes the word and σ the syllables.

as ‘working’ consists of the constituents ‘work’ and ‘ing’. The first syllable, ‘work’ has an onset(‘w’), nucleus(‘o’) and complex coda (‘rk’), the second constituent however has only a nucleus (‘i’) and a coda (‘ng’). See Figure 1 for a graphical representation of the constituents of ‘working’.

Phonotactic constraints define permissible,

language dependent phoneme sequences. For example, the velanised nasal (/ŋ/) can occur in coda position in English, e.g. ‘thing’, but never as an onset. Other languages however may have different constraints, e.g. Vietnamese allow /ŋ/ as an onset. These differences among languages seem to indicate that phonotactic patterns must be learned (Moreton 2002, Wilson 2010).

We can distinguish first and second order constraints. First order constraints are only dependent on the position of the constituent in the syllable. Second order constraints also take into account other features of the syllable, such as the onset and the coda. Thus, never allowing the /ŋ/ as the onset is a first order constraint, while an example of a second order constraint would be that /ŋ/ may only be the coda when /i/ is the nucleus. Phonotactic constraints usually apply to natural classes of sounds, e.g. in Dutch a phonotactic constraint is that codas must always be unvoiced.

1.1 Phonotactic Learning

When learning phonotactic regularities speech sound representations must be abstracted across contexts. This is for example needed to filter out the effect that nearby sounds have on the pronunciation. However other context-dependent differences may be the results of phonotactic constraints. Therefore these constraints can not be learned without context (Kessler & Treiman 1997). From this follows that the representations of speech sounds must be flexible for children to be able to learn phonotactic constraints.

De Boysson-Bardies & Vihman (1991) have shown that nine-month old children have detailed knowledge of the phonotactic regularities of their mother tongue, as shown

in their preference for certain syllables when babbling. The children in this experiment did not start speaking until they were about 17 months old. This result illustrates that one can acquire sensitivity for phonotactic constraints by simply listening.

That both adults and children have flexible representations of spoken words has been shown by Fisher et al. (2001). The representations of spoken words are updated with both abstract information and token-specific details after each listening experience. This indicates that the learning mechanisms that are present in young children are still existent in adults. If the representation of a word changes after hearing it, then phonotactic knowledge could be accumulated by repeated auditory experience.

Finally Onishi et al. (2002) demonstrated that phonotactic constraints are rapidly learned from listening experience. Adults required as little as five auditory presentations of each token in an artificial language, created from ten consonants and two vowels, to acquire the phonotactic constraints of that language. This was true for both first and second order constraints. Knowledge of the restrictions was tested by measuring the difference in reaction time between repeating a word after hearing it. Reaction times were lower for words that conformed to the phonotactic constraints than for words that violated the restrictions.

The above research thus established that both adults and children have flexible representations of words and that both can learn phonotactic constraints rapidly by just listening.

1.2 Zipfian Learning

It has been shown that the learning of categories is optimized by the use of a low-variance sample centred on prototypical exemplars (Elio & Anderson 1981, 1984). Such low-variance samples allow subjects to easily find the common factor which can then be generalized.

The Zipfian distribution generates such low-variance samples. This distribution has been shown to approximate the distribution of word types in natural language. (Zipfian 1935). Compare the frequency of the definite article ‘the’ to the frequency of far less used ‘offal’, which is defined as the entrails and internal organs of an animal used as food.

The acquisition of phonotactic constraints is equivalent to the learning of categories. One learns two categories of phonemes, the group of consonants that can be onset and the consonants that are permissible in the coda position.

Second order constraints can be seen as the learning of categories by seeing the coda and the onset as a frame in which the nucleus is placed. These frames can be categorized based on the nuclei they accept.

Based on this view of the learning of phonotactic constraints and the above mentioned research it is probable that learning phonotactic constraints in the Zipfian distribution has a better result than learning in the uniform distribution.

This idea is further supported by two experiments done by Kurumada et al. (2011). They compared the uniform and Zipfian distribution in experiments on word recognition and word segmentation of an artificial language/ The artificial language they used contained words built from two to four different syllables, which were randomly concatenated into sentences of four words. The artificial languages contained no phonotactic constraints.

In the word recognition task subjects listened to a sample of speech of one of four artificial languages. The languages had a different number of types, namely 6, 12, 24 or 36 words. The used language contained 300 different words, which means that each type was presented 50, 25, 12.6 or 8.7 times on average in the uniform condition. Subjects were tested in a two-alternative forced-choice task in which they were asked what word sounded more like it came from the language they had just listened to, performance was measured as the number of correct answers. Performance was affected by the frequency of the types, independent of the distribution.

In the word segmentation task subjects listened to one of four languages that were comparable with those used in the word recognition task. Languages contained 6, 9, 12 or 24 word types. Since each subject saw 200 tokens in 50 sentences, types were presented 33.3, 22.2, 16.7 or 8.3 times on average in the uniform condition. Subjects were asked to indicate word boundaries in visually presented transcriptions, after hearing a synthesized sentence as many times as needed. Subjects were scored on how well they segmented words. Participants who were exposed to the Zipfian distributions generally achieved higher performance than those exposed to the uniform distribution. Only when the language contained 6 types was the performance comparable.

The results showed that adult learners performed significantly better on a segmentation task when trained in the Zipfian distribution. This indicates that the Zipfian distribution allows listeners to learn the high-frequency words and use them to better handle the less known words, whereas those who learned an uniform language can not apply such bootstrapping. However the uniform distribution and the Zipfian distribution achieved comparable performance in the word recognition task.

To summarize, previous literature leaves us with the impression that the Zipfian distribution might help in learning phonotactic constraints.

1.3 The Present Research

The hypothesis guiding this research is that second order phonotactic constraints are better learned when they are presented in a Zipfian distribution than in a uniform distribution.

Subjects listened to consonant-vowel-consonant (CVC) words in one of two artificial languages. These languages were defined using two sets of consonants and two nuclei. The position of the consonants was dependent on the vowel, between the two nuclei the consonant positions were reversed. Thus if the first vowel, e.g. /æ/, is associated with the frame b_p then the other vowel, e.g. /ɪ/, is used with the frame p_b.

These CVC words were presented in either the uniform or the Zipfian distribution. All languages were split in a training and a test set of equal size, which were balanced across nucleus, onset, coda and lexical status of the words. Learning was assessed by presenting subjects with all words of both languages and asking them to repeat the words as quickly as possible without making errors. If phonotactic constraints are better learned with a Zipfian distribution, subjects trained in this distribution should score higher during the test phase than those trained in the uniform distribution.

2 Method

2.1 Subjects

18 native speakers of Dutch (5 women, mean age = 22 years, range = 19-32) who reported normal hearing took part in the experiment. The subjects received a cash payment of five

euros for their participation. No subjects were excluded. 6 participants were trained in the uniform distribution with Language 1, all other conditions had 4 participants.

2.2 Stimuli

Two groups of consonants based on Onishi et al. (2002) were used in this experiment (Group 1: /b, m, ð, k/; Group 2: /p, t, n, g/). These consonants cannot be distinguished by a small set of phonetic features. They were combined with the nuclei (/æ/, /ɪ/). These vowels were chosen because they are easily discriminable although the choice for a certain vowel does not seem to influence the generalization of phonotactic constraints (Chambers et al. 2010).

These consonants and vowels were combined to form an artificial language in which the consonant positions depended on the adjacent vowels; thus /b/'s might be onset and /p/'s codas with the vowel /æ/, but these positions would be reversed with the vowel /ɪ/ (Onishi et al. 2002). A female native speaker of English recorded the syllables. The same tokens were used in study and test.

Each of these two lists were equally divided in a test set and a set of training tokens. These sets were balanced concerning the frequency of the onset, nucleus and coda. Furthermore the words with a lexical status and without it were in the same proportion in both sets. The training and test sets of both languages are listed in Appendix A. The test set for all four conditions consisted of the training and test tokens of both languages. For the subjects in one of conditions of Language 1 all items of Language 2 were illegal and those of Language 1 legal. This was the other way around for subjects trained in Language 2.

Emotional words, e.g. vulgar words such as

‘bitch’ were not removed, but only admitted to the test set, to avoid removing information while still nullifying the effect of the better remembrance of emotional words (Kensinger & Corkin 2003). In Language 1, 17 of the 32 words had a lexical status in either Dutch or English, in Language 2 this was the case for 11 of the 32 words.

As shown previously the difference in the number of words with a lexical status does not interact with performance when words and non-words are intermixed and no lexical access is required (Onishi et al. 2002, Mirman et al. 2008).

2.3 Procedure

Subjects were randomly assigned to one of the four conditions. The syllables on the training lists were presented in a random order for each subject. During the training phase subjects listened to their training lists and were asked to rate the pronunciation of the word on a scale of 1 to 5. After a distraction task subjects listened to the test set and repeated the items as quickly as possible without making errors. The full experiment was recorded using the built in microphone of the laptop used for testing.

2.4 Scoring

The recordings of the repetition test were transcribed. Errors were excluded from further analysis ($\mu = 6.9$ per subject). Words that rated lower than a 3 on understandability were removed (Language 1: ‘nit’, Language 2: ‘nat’). Reaction times were measured from stimulus offset to response onset. Responses more than 1.5 s after stimulus offset were removed (1 item) as were responses standard deviations beyond each subjects mean ($\mu = 2$ responses per subject, $\sigma = 1.0$).

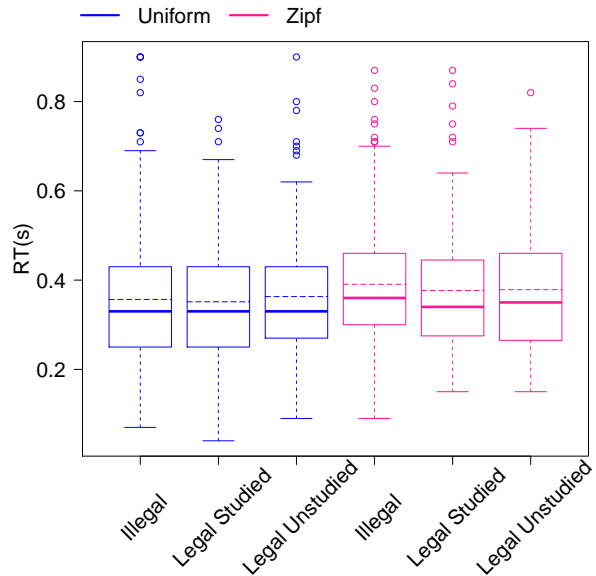


Figure 2: A box plot of the data. The solid line denotes the median, while the dashed line represents the mean. The whiskers extend from $Q1 - 1.5 \cdot QRQ$ to $Q3 + 1.5 \cdot QRQ$.

3 Results

In the section below two different analyses of the data are presented. An overview of the data can be found in an overview in Figure 2. In this plot the median and average reaction time are shown per type of item (legal studied, legal unstudied, illegal) separated per distribution. The plot illustrates that there is some difference in mean between the two distributions, the subjects trained in the Zipfian distribution seem to perform worse than those in the uniform distribution.

3.1 Classical Analysis

A two-way analysis of variance by subject yielded no main effects for type ($F_1(2, 45) = 0.001$, $p > 0.999$). Neither the factor distribution ($F_1(2, 45) = 0.185$, $p = 0.669$) nor the interaction effect was significant ($F_1(2, 45) = 0.018$, $p > 0.982$). Indicating that type nor

distribution had a significant effect on the response time.

In a two-way analysis of variance by item distribution was a main effect ($F_2(2, 227) = 6.354, p = 0.0124$) such that the average response time was significantly faster for the Zipfian distribution ($\mu = 0.38, \sigma = 0.078$) than for the uniform distribution ($\mu = 0.36, \sigma = 0.061$). Both the effect of type ($F_2(2, 227) = 0.745, p = 0.4757$) and the interaction effect ($F_2(2, 227) = 0.727, p = 0.4846$) were insignificant.

Onishi et al. (2002) found that the advantage of legal over illegal items decreased during testing since repetition of later words is influenced by listening to words exhibiting second order constraints. To investigate whether or not we have a comparable effect we dived de data in two equal blocks, see Figure 3 which shows the decreasing advantage of legality. Subjects did not have a significant lower reaction time on legal items in block 2 compared to block 1 ($F(1, 17) = 1.1414, p = 0.3941$). There was no significant legality advantage in either block 1 ($F(1, 17) = 1.1396, p = 0.6047$) or block 2 ($F(1, 17) = 0.8931, p = 0.4092$). The difference between studied legal words and unstudied legal words was not significant in block 1 ($F(1, 17) = 1.1416, p = 0.606$) nor in block 2 ($F(1, 17) = 1.1482, p = 0.6105$). The effects of the number of preceding items in the test phase that has been found in a comparable experiment by Onishi et al. could not be established in this experiment.

3.2 Linear Mixed Effect Models

To avoid the language-as-a-fixed-effect fallacy (Clark 1973) the data were also analysed using Linear mixed effect models with subjects and items as random effects (Baayen et al. 2008). Visual inspections of plots of residuals against fitted values were used to check for

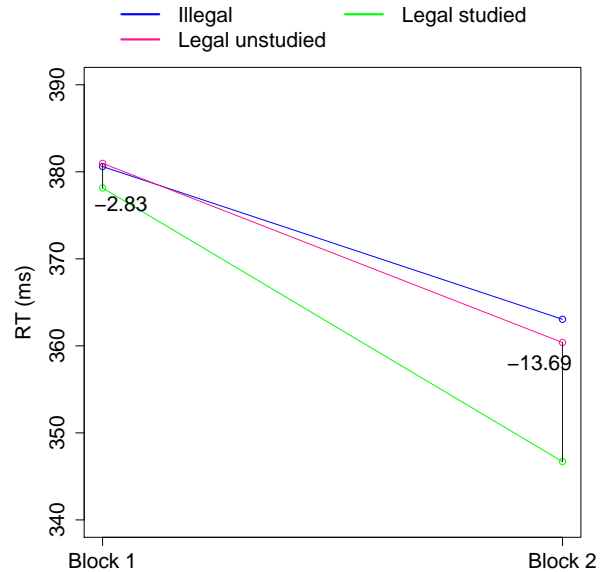


Figure 3: Results from the experiment dived in two blocks of trials.

normality and homogeneity. A likelihood ratio test was used to compare the models with the fixed effects to null models with only the random effects. Results that did not differ significantly from the null model were rejected. The p-values presented in this section are MCMC-estimated.

The fixed effects of the experiment and thus of the model were distribution and type. No model with item and subject as a random effect showed a significant influence from distribution and type on the reaction time, which is consistent with the classical analysis presented in section 3.1.

3.3 Analysis of the Uniform Group

Onishi et al. found that legal words were repeated significantly faster than illegal ones. In our data no such difference ($t(9) = -0.2556, p = 0.804$) was found. Contrary to Onishi et al. we did not find a reliable decrease in the advantage of legality between the first and second half of the test items ($F(1,9) = 1.0778,$

$p = 0.4565$).

4 Discussion

This section will aim to present some possible explanations for the lack of significant results and some suggestions for further research will be made.

4.1 Experimental Set-up

The group trained in the Zipfian distribution did not perform reliably better than those trained in the uniform condition according to the item analysis. As has been shown in section 3.3 the effects found by Onishi et al. have not been replicated in this experiment. The lack of any significant results may have been due to the set-up of the experiment.

The described experiment differed on some points from that done by Onishi et al. Due to time and resource constraints the number of subjects was much lower (16 versus 40).

Another difference when compared to Onishi et al. is that the used languages only contained 32 tokens. This reduced the number of items subjects could use to generalize with.

Not only were the languages smaller, the number of presentations per token was also lower. Onishi et al. used an average of five presentations per word, while we had an average of four.

In conclusion it is possible that five is the minimal number of presentations necessary to learn phonotactic constraints, meaning that our four presentations were too few to allow phonotactic regularities to be acquired.

Kurumada et al. (2011) found that the Zipfian distribution results in higher performance in word segmentation but not in word recognition.

Compared to the word recognition task of Kurumada et al. we used a lower number of

average presentations (4) compared to the at least 8.3 presentations. Another difference lies in the paradigm, we tested whether or not the word was known by measuring reaction time when repeating words, Kurumada et al. explicitly asked for this information, by asking subjects if they recognised a word. In spite of the paradigm differences the results were the same: distribution was not a significant factor.

As stated earlier, in learning phonotactic constraints subjects have to segment the words to discover the phonotactic constraints. Kurumada et al. found that the Zipfian distribution offers an advantage in segmentation when the lexicon contains nine or more words. However the segments of the words used in our experiment are uniformly distributed. It may thus be possible that distributing the constituents used to form the words according to the Zipfian distribution aids the segmentation of words and thus influences learning phonotactic constraints.

Distributing the words according to the Zipfian distribution however, did not aid in learning phonotactic constraints, when the average number of presentations is less than four, according to our experiment.

In conclusion it is probable that one needs at least five auditory presentations when one tries to learn phonotactic constraints.

Distributing the syllables instead of the words according to the Zipfian distribution might aid in the learning of phonotactic constraints.

4.2 Distribution

The Yule distribution is a more general form of the Zipfian distribution and is generally used to describe the distribution of the different codons in DNA (Yule 1925). Tambovtsev & Martindale (2007) have shown that the Yule distribution is a better fit for the phoneme distribution

of English and Dutch, among others, than the Zipfian distribution.

It is possible that the cognitive systems that guide the learning of regularities of phonemes, such as phonotactic constraints, would respond better to the Yule distribution. Thus phonotactic constraints might be better learned when the constituents are distributed according to the Yule distribution, compared to the uniform distribution.

4.3 Natural Classes

The artificial language we used was more complex than the language used by Onishi et al.. While they used three different kinds of plosives: bilabial, alveolar and velar, we used four, adding a retroflex plosive.

Another difference is that we used a dental fricative while they used an alveolar fricative.

Thus Onishi et al. used fewer natural classes than we did, making their language less complex than ours. Since we tried to avoid using multiple phonemes from the same natural class our language contained more natural classes. The higher number of natural classes might have contributed to the lack of significant results, as they made the language more complex and thus possibly harder to learn.

4.4 Conclusion

In this section it has been explained that the lack of significant results might be due to too few presentations during the training phase. It appears thus that we may have hit a minimal number of auditory presentations necessary for phonotactic learning. At least five presentations seem necessary for learning when using an uniform distribution.

Another possible explanation is the use of an additional phoneme from another natural class. This might have made the artificial language too complex to learn the phonotactic

constraints with such low number of presentations.

It has furthermore been suggested that distributing phonemes according to the Zipfian distribution or the Yule distribution might achieve better results than the uniform distribution in learning phonotactic constraints.

When trying to determine the benefits of the Zipfian distribution in phonotactic learning one should at least use five tokens per type in the uniform distribution.

References

- Baayen, R., Davidson, D. & Bates, D. (2008), 'Mixed-effects modeling with crossed random effects for subjects and items', *Journal of memory and language* **59**(4), 390–412.
- Chambers, K., Onishi, K. & Fisher, C. (2010), 'A vowel is a vowel: generalizing newly learned phonotactic constraints to new contexts.', *Journal of Experimental Psychology: Learning, Memory, and Cognition* **36**(3), 821–828.
- Clark, H. (1973), 'The language-as-fixed-effect fallacy: A critique of language statistics in psychological research', *Journal of verbal learning and verbal behavior* **12**(4), 335–359.
- De Boysson-Bardies, B. & Vihman, M. M. (1991), 'Adaptation to language: Evidence from babbling and first words in four languages', *Language* **67**(2), 297–319.
- Elio, R. & Anderson, J. (1981), 'The effects of category generalizations and instance similarity on schema abstraction.', *Journal of Experimental Psychology: Human Learning and Memory* **7**(6), 397–417.
- Elio, R. & Anderson, J. (1984), 'The effects of information order and learning mode on

- schema abstraction', *Memory & cognition* **12**(1), 20–30.
- Fisher, C., Hunt, C., Chambers, K. & Church, B. (2001), 'Abstraction and specificity in preschoolers' representations of novel spoken words', *Journal of Memory and Language* **45**(4), 665–687.
- Fromkin, V., Curtiss, S., Hayes, B., Hyams, N., Keating, P., Koopman, H., Munro, P., Sportiche, D., Stabler, E., Steriade, D. et al. (2000), *Linguistics: An introduction to linguistic theory*, Blackwell, chapter 13.
- Kensinger, E. & Corkin, S. (2003), 'Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words?', *Memory & cognition* **31**(8), 1169–1180.
- Kessler, B. & Treiman, R. (1997), 'Syllable structure and the distribution of phonemes in english syllables', *Journal of Memory and Language* **37**(3), 295–311.
- Kurumada, C., Meylan, S. C. & Frank, M. C. (2011), Zipfian word frequencies support statistical word segmentation, in 'Cognitive Science Conference'.
- Mirman, D., McClelland, J., Holt, L. & Magnuson, J. (2008), 'Effects of attention on the strength of lexical influences on speech perception: Behavioral experiments and computational mechanisms', *Cognitive science* **32**(2), 398–417.
- Moreton, E. (2002), 'Structural constraints in the perception of english stop-sonorant clusters', *Cognition* **84**(1), 55–71.
- Onishi, K., Chambers, K. & Fisher, C. (2002), 'Learning phonotactic constraints from brief auditory experience', *Cognition* **83**(1), 13–23.
- Tambovtsev, Y. & Martindale, C. (2007), 'Phoneme frequencies follow a yule distribution', *SKASE Journal of Theoretical Linguistics* **4**(2), 1–11.
- Wilson, C. (2010), 'Learning phonology with substantive bias: An experimental and computational study of velar palatalization', *Cognitive science* **30**(5), 945–982.
- Yule, G. U. (1925), 'A mathematical theory of evolution, based on the conclusions of dr. jc willis, frs', *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character* **213**, 21–87.
- Zipfian, G. (1935), *The psycho-biology of language*, Houghton, Mifflin.

A Stimuli

Table 1: *The syllables are formed using consonants from Group 1 and 2 and the nuclei. A zero indicates a non-lexical status and 1 a lexical status, in the training column indicates a 1 that the word has been used while training and a 0 that a word has not been used during training.*

ID	Word	Lexical status	Language	Training
1	bap	0	1	0
2	bag	1	1	1
3	ban	1	1	1
4	bach	1	1	0
5	bip	0	2	1
6	big	1	2	1
7	bin	1	2	0
8	bich	1	2	0
9	kap	1	1	1
10	kag	0	1	1
11	kan	1	1	0
12	kach	1	1	0
13	kip	1	2	0
14	kig	0	2	1
15	kin	1	2	1
16	kich	0	2	0
17	map	1	1	1
18	mag	0	1	0
19	man	1	1	0
20	mach	1	1	1
21	mip	0	2	0
22	mig	1	2	0
23	min	1	2	1
24	mich	0	2	1
25	tap	1	1	0
26	tag	1	1	0
27	tan	1	1	1
28	tach	0	1	1
29	tip	0	2	1
30	tig	0	2	0
31	tin	0	2	0
32	tich	0	2	1
33	pab	0	2	1
34	pak	1	2	1
35	pam	0	2	0

Table 1: *The table from page 10 continued.*

ID	Word	Lexical status	Language	Training
36	pat	1	2	0
37	pib	0	1	0
38	pik	1	1	1
39	pim	0	1	1
40	pit	1	1	0
41	gab	0	2	0
42	gak	0	2	1
43	gam	0	2	1
44	gat	0	2	0
45	gib	0	1	1
46	gik	0	1	1
47	gim	0	1	0
48	git	1	1	0
49	nab	0	2	0
50	nak	0	2	0
51	nam	0	2	1
52	nat	0	2	1
53	nib	0	1	1
54	nik	0	1	0
55	nim	0	1	0
56	nit	0	1	1
57	chab	0	2	1
58	chak	1	2	0
59	cham	0	2	0
60	chat	1	2	1
61	chim	0	1	0
62	chik	1	1	0
63	chit	1	1	1
64	chib	0	1	1