ON THE PROCESSING OF NONWORDS IN WORD NAMING AND AUDITORY LEXICAL DECISION

Yu-Ying Chuang[1], Marie-lenka Vollmer[1], Elnaz Shafaei-Bajestan[1],
Susanne Gahl[2], Peter Hendrix[1], and R. Harald Baayen[1]

Eberhard-Karls University of Tübingen[1]; University of California at Berkeley[2]
yu-ying.chuang@uni-tuebingen.de; marie.voller@online.de; elnaz.shafaei-bajestan@uni-tuebingen.de;
gahl@berkeley.edu; peter.hendrix@gmail.com; harald.baayen@uni-tuebingen.de

ABSTRACT

Nonwords are often used to clarify how lexical processing takes place in the absence of semantics. This study shows that nonwords are not semantically vacuous. We used Linear Discriminative Learning [2] to estimate the meanings of nonwords in the MALD database [14] from the speech signal. We show that measures gauging nonword semantics significantly improve model fit for both acoustic durations and RTs. Although nonwords do not evoke meanings that afford conscious reflexion, they do make contact with the semantic space, and the angles and distances of nonwords with respect to actual words co-determine articulation and lexicality decisions.

Keywords: Nonword, auditory lexical decision, lexical processing, Linear Discriminative Learning.

1. INTRODUCTION

Nonwords are widely used in a variety of linguistic experiments. In lexical decision tasks, for example, they are used not only to balance word decisions with nonword decisions, but also to help conceal the experimental manipulation of word materials.

When nonwords are themselves the target of study, the obvious absence of clear meanings for nonwords has led researchers to use nonwords to study lexical processing without any interference of semantics. For example, Vitevitch and Luce [15] contrasted nonwords with words to distinguish the loci where different effects occur. Examining the effects of phonological neighborhood density and phonotactic probability on spoken word recognition with a shadowing task, they found effects with opposite sign for words and nonwords. Whereas higher diphone probability and denser similarity neighborhoods predicted longer reaction times for words, they predicted shorter reaction times for nonwords.

Vitevitch and Luce argued that the effects of phonological neighborhood density and phonotactic probability arise at different processing levels.

The effect of neighborhood density was interpreted as reflecting lexical competition, the hallmark of word processing. The effect of phonotactic probability, which in their data was strongly correlated with neighborhood density, was understood as arising at the sublexical level. Nonword processing would then be driven primarily by phonotactic probability.

In a subsequent study, Vitevitch and Luce [16] reported that in lexical decision, nonword processing can also involve the lexical level, if nonwords co-activate real words that then enter into a process of lexical competition. However, by assumption, nonwords still do not have their own specific semantics.

In this study we show that nonwords are not semantically empty and that nonword-specific meanings do co-determine lexical processing. This idea is in fact not new. Take sound symbolism, for example. If nonwords would not activate any semantics, it is hard to explain the well-established finding that certain nonwords are consistently associated with specific semantic concepts: a nonword such as *baluma* is typically associated with objects with round curvy shapes ([9, 13]).

The goal of the present study is to complement investigations of phonological aspects of nonword processing with an exploration of their semantic effects. To that end, we extracted the nonword data from the Massive Auditory Lexical Decision (MALD) database [14], and examined acoustic durations and reaction times. Specifically, we generated semantic vectors for nonwords, using the algorithms of a new cognitively and linguistically motivated theory of the mental lexicon, Linear Discriminative Learning [2].

2. METHOD

2.1. Materials

We obtained the recordings and reaction time data of the nonwords from MALD. The nonwords, all composed of phonotactically legal syllables, were recorded by one male native speaker of English and

were responded to by 231 native listeners in a lexical decision task. In what follows, we analyzed only correct responses. This left us with 9,573 data points for the acoustic durations, and 96,049 for the RTs.

2.2. Estimating nonword meaning

Following Landauer and Dumais [11] and subsequent work in distributional semantics, we approximated word meanings by means of vectors of real-valued numbers, henceforth semantic vectors. We made use of the semantic vectors described in Baayen et al. [2], which were derived from the TASA corpus [8, 12]. The computational model of Linear Discriminative Learning (henceforth LDL) developed in this study maps a numeric vector representing a word's form onto a numeric vector representing a word's meaning.

Formally, let C_w denote the matrix specifying words' forms, and let S_w denote the matrix specifying words' meanings. The form and meaning matrices have the same number of rows, one for each word. Thus, in C_w, a given row vector c_w represents the form features of one specific word, whereas in S_w, the corresponding row represents the semantic vector (s_w) of that word. We define a matrix F that maps the form vectors in C_w as precisely as possible onto the semantic vectors of S_w. Using the notation of linear algebra, this mapping can be written as:

$$(1) \quad C_w F = S_w.$$

The mapping F is equivalent to a two-layer network without hidden layers. Given the matrices C_w and F, equation (1) can be solved (see [2] for mathematical details). Given the mapping F and a set of nonwords, the form vectors of which are brought together as the row vectors of a matrix C_{nw}, we can estimate the semantic vectors of these nonwords by post-multiplying the matrix of form vectors with F:

$$(2) \quad C_{nw} F = \hat{S}_{nw}.$$

Form and meaning vectors were constructed for the 19,412 words from MALD for which TASA-based semantic vectors were also available to us (cf. [2]). The semantic vectors constructed in this study were used to build the semantic matrix S_w. The form matrix C_w was derived from the speech signal using the Frequency Band Summary (FBS) features [1]. For the present dataset, there was a total of 26,336 distinct FBS features. A word's form vector specifies which of these FBS features are present in that word, using binary coding with 1 for presence and 0 for absence. For the 9,573 nonwords in MALD, we created a second form matrix (C_{nw}), following the same

procedure as for the word matrix C_w. Using (2), we then derived the semantic matrix for nonwords $\hat{S}_{nw}$, each row of which represents the predicted semantic vector ($\hat{s}_{nw}$) of a nonword.

2.3. Predictors

For the analysis we considered five variables as predictors. The first two are measures of phonological form, whereas the latter three are semantic measures derived from $\hat{S}_{nw}$.

Phonological Neighborhood Density (PhonND): The number of words which have an edit distance of one phone (by addition, deletion, or substitution) from the nonword. This measure is provided by MALD.

Biphone Phonotactic Probability (BiphProb): The mean biphone phonotactic probability of a nonword. This measure was obtained from the online phonotactic probability calculator developed by Vitevitch and Luce [17].

Euclidean Distance from Nearest Neighbor (EDNN): The Euclidean distance of the position of a nonword in the semantic space from the position of its nearest word neighbor. This was calculated by using the FNN package of R [3].

Average Lexical Correlation (ALC): The mean of the correlations of a nonword's semantic vector $\hat{s}_{nw}$ with each of the words' semantic vectors s_w. Correlation, similar to consine similarity, is a measure of the angle of two vectors. The smaller the angle is, the more similar the two vectors are, and hence the higher the correlation. Higher ALC indicates that the nonword has "landed" in a denser semantic neighborhood.

Nearest Neighbor Correlation (NNC): The maximum correlation between a nonword semantic vector $\hat{s}_{nw}$ and a word semantic vector s_w, taking into account all word vectors. Its value is high when the angle between $\hat{s}_{nw}$ and s_w is small, suggesting the nonword is semantically similar to a word.

3. RESULTS

3.1. Nonword duration

Statistical analyses were conducted using the Generalized Additive Mixed Model (GAMM) [18], which allows the modeling of nonlinear functional relations between one or more predictors and the response variable. To avoid problems of interpretation due to multicollinearity (the pairwise correlations of these measures are: EDNN-ALC: $r = -0.28$; EDNN-NNC: $r = -0.45$; ALC-NNC: $r = 0.57$), we performed a Principal Component Analysis (PCA) on

the three semantic measures. The PCA loadings for these measures are presented in Table 1. We included PC1 and PC2, which together accounted for 87% of the variance in the semantic measures, into our regression analyses.

We fitted a GAMM to the nonword durations, with as predictors PC1 and PC2, as well as from the two phonological measures of `PhonND` and `BiphProb`. The upper panels of Figure 1 visualize the partial effects of the two phonological factors. Both are well-supported predictors for nonword duration, but their effects are opposite. Accordingly, nonwords in a denser phonological neighborhood are produced with shorter duration, while nonwords with higher phonotactic probability have longer duration. This pattern of results fits well with previous work on `PhonND` effects on duration. Gahl et al. [7], based on a different model, the two-step interactive activation model of lexical access of Dell [5] and Dell and Gordon [6], hypothesized that increasing `PhonND` should be associated with shorter word durations. That prediction was confirmed based on a corpus of spontaneous speech.

The semantic measures, the partial effects of which are presented in the bottom panels of Figure 1, are also predictive. Inclusion of these predictors increased model fit substantially (Δ AIC = 1845 units). For both PCs, we found upward-sloping trends. Since PC1 is a dimension that contrasts `EDNN` on the one hand with `ALC` and `NNC` on the other hand, if PC1 is large, a nonword is distant from other words, has landed in a sparse neighborhood, and is semantically dissimilar to any word. Under this condition, durations increase. PC2 is an orthogonalized dimension that aligns `EDNN` and `ALC`. A large value of PC2 indicates that the nonword is again distant from other words, but now it is in a dense region of semantic space. In this case durations increase as well. PC1 may be capturing nonwords that stand out semantically as nonwords, similar to the way in which a novel taste can stand out as distinct from previously experienced tastes. The lengthening effect of PC1 may thus relate to previous work on spoken word duration [10, 4] that observed longer word durations in connection with decreased entropy. PC2, on the other hand, may reflect what in interactive activation models would be understood as competition among semantic neighbors [6], resulting in longer durations in speech production.

3.2. Nonword RTs

Most responses were executed after nonword offset, with only a few exceptions (1.5%). Given that nonword duration is well correlated with reaction

Table 1: PCA loadings for the semantic measures.

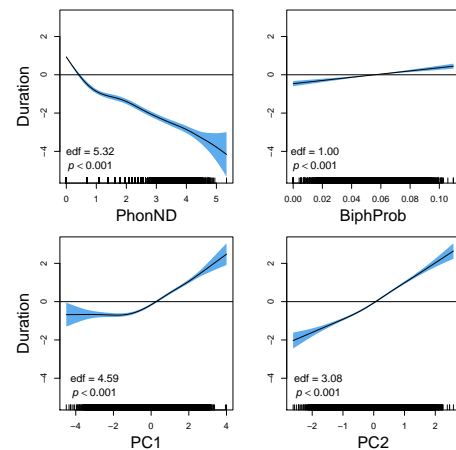|	PC1	PC2	PC3
EDNN	0.51	0.80	-0.30
ALC	-0.58	0.59	0.58
NNC	-0.64	0.12	-0.76

Figure 1: Predictor effects on nonword duration. The shaded area indicates 2 standard error bounds.

time ($r = 0.37$), and is the dominant predictor when reaction times are measured from word onset, we focused on reaction times measured from nonword offset. Responses that were made before stimulus offset were excluded. RTs were Box-Cox transformed ($\lambda = 0.26$).
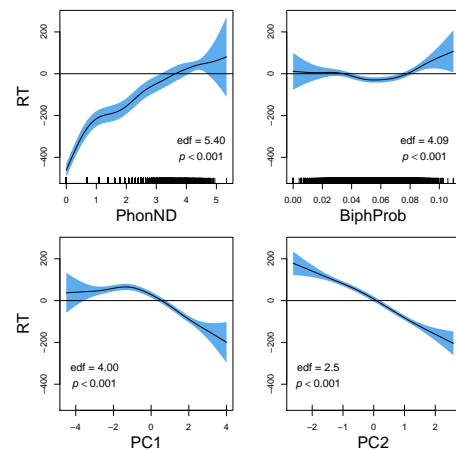

Figure 2: Predictor effects on nonword RTs. The shaded area indicates 2 standard error bounds.

We again made use of a GAMM to predict RTs from the two phonological measures and the two semantic PCs, now also including by-subject and by-nonword random intercepts. Compared to a baseline model with only the two phonological measures,

model fit significantly improved with the inclusion of the two semantic measures (Δ AIC = 141 units).

As shown in the upper left panel of Figure2, RTs increase with `PhonND`, suggesting that it is harder to reject a nonword when it has many word neighbors of similar phonological make-up. The effect of `BiphProb`, on the other hand, is much attenuated and irregular in shape. Both semantic measures reveal downward trends. As for duration, the effect of PC1 levels off for low values.

4. DISCUSSION

This study shows that nonword processing is influenced not only by form similarity (as gauged by phonological neighborhood density and biphone phonotactic probability) but also by nonword semantics. First consider the effects of form similarity.

Form effects on RTs support Vitevitch and Luce's proposal that nonword processing in the lexical decision task is determined by `PhonND` [16]: The effect of `PhonND` on RTs was strong, whereas that of `BiphProb` was weak and U-shaped. For words, Gahl et al. [7] showed that `PhonND` and `BiphPhon` are positively and strongly correlated, and have similar effects on acoustic duration. For the nonwords of our study, however, their effects are opposite, and the two measures are not well correlated ($r = -0.12$). This may be due to the much wider range of syllable types and word lengths in the MALD nonword data. For instance, when we restrict the dataset to CVC nonwords, the syllable type figuring in Gahl et al., a positive correlation between `PhonND` and `BiphProb` emerges. Another potential cause that leads to the discrepancy between the present study and the previous one is genre difference. While we examined productions elicited by a naming task, Gahl et al. investigated spontaneous speech. We leave the interaction of syllable structure, word length, morphological complexity, genre type, and phonological measures for future investigation, as the focus of the present study is on semantic influences on RTs and durations.

Figure 3 helps clarify the semantic effects we observed. The semantic vector of a target nonword is shown in blue, and the vectors of two words are shown in red and orange. These words are at the same Euclidean distance d from the target, but the angles between their vectors and the target nonword vector differ: $\alpha_1 > \alpha_2$. The EDNN measure is the radius d to the closest neighbor. The `ALC` and `NNC` measures relate to the angles α. The `ALC` measures how closely oriented the nonword target vector is to the mean of all word vectors, and the `NNC` measures
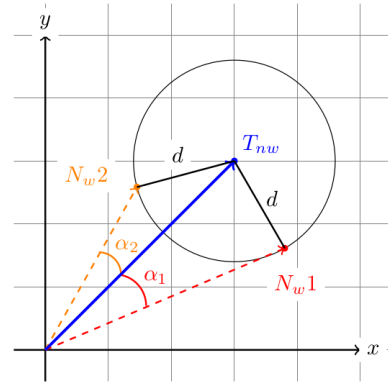

Figure 3: Semantic neighbors (N_w1 and N_w2) can be at the same distance d of a target nonword T_{nw} and yet be at different angles (α_1 and α_2) from the target vector, and hence differ in their semantic similarity (as guaged with the cosine similarity or correlation measures).

the similarity in orientation for the neighbor with the smallest angle α. Larger positive values of PC1 represent nonwords with a large radius d and large angle α_1, i.e., less semantically similar neighbors. Orthogonal to this, larger values on PC2 represent nonwords with large radius d but small angles (α_2), i.e, semantically more similar words.

The analysis of the RTs suggests that the radius d may inform lexicality decision, as for nonwords with larger radius d (i.e., larger EDNN), reaction times are shorter. Such stimuli are clearly nonwords. The leveling off of the effect of PC1 on RTs and durations may reflect the consequences of using a threshold θ for nonword decisions based on the radius d:

$$\text{RT} \propto \text{I}_{[d \le \theta]} a + \text{I}_{[d > \theta]} (a - bd),\ b > 0,$$

where $\text{I}_{[x]}$ evaluates to 1 if x is true, and to 0 otherwise. For stimuli with large d, their nonword status is so clear that they can be produced with confidence by the speaker reading out the nonwords, resulting in longer durations [10, 4].

In conclusion, the present study offers evidence that nonwords do project into semantic space, and that where they land in this space affects both speech production (acoustic durations) and comprehension (RTs). Furthermore, it appears that not only semantic similarity (the angles between nonword and word vectors) but also the Euclidean distance between these vectors plays a substantial role. Possibly, lexicality decisions are based, at least in part, on whether the semantic distance of a nonword to its nearest word neighbor is greater than some fixed threshold.

5. REFERENCES

[1] Arnold, D., Tomaschek, F., Lopez, F., Sering, T., Baayen, R. H. 2017. Words from spontaneous conversational speech can be recognized with human-like accuracy by an error-driven learning algorithm that discriminates between meanings straight from smart acoustic features, bypassing the phoneme as recognition unit. *PLOS ONE* 12(4), e0174623.

[2] Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., Blevins, J. P. 2019. The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity* 1–39.

[3] Beygelzimer, A., Kakadet, S., Langford, J., Arya, S., Mount, D., Li, S. 2018. *FNN: Fast Nearest Neighbor Search Algorithms and Applications*. R package version 1.1.2.1.

[4] Cohen, C. 2014. Probabilistic reduction and probabilistic enhancement. *Morphology* 24(4), 291–323.

[5] Dell, G. S. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological review* 93(3), 283.

[6] Dell, G. S., Gordon, J. K. 2003. Neighbors in the lexicon: Friends or foes? In: Schiller, N. O., Meyer, A. S., (eds), *Phonetics and phonology in language comprehension and production: Differences and similarities* volume 6. Berlin: Mouton de Gruyter 9–37.

[7] Gahl, S., Yao, Y., Johnson, K. 2012. Why reduce? phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66(4), 789–806.

[8] Ivens, S. H., Koslin, B. L. 1991. *Demands for Reading Literacy Require New Accountability Methods*. Touchstone Applied Science Associates.

[9] Köhler, W. 1929. *Gestalt psychology*. New York: Liveright.

[10] Kuperman, V., Pluymaekers, M., Ernestus, M., Baayen, R. H. 2006. Morphological predictability and acoustic salience of interfixes in Dutch compounds. *JASA* 122, 2018–2024.

[11] Landauer, T., Dumais, S. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review* 104(2), 211–240.

[12] Landauer, T. K., Foltz, P. W., Laham, D. 1998. Introduction to latent semantic analysis. *Discourse Processes* 25, 259–284.

[13] Sapir, E. 1929. A study in phonetic symbolism. *Journal of Experimental Psychology* 12(3), 225–239.

[14] Tucker, B. V., Brenner, D., Danielson, D. K., Kelley, M. C., Nenadić, F., Sims, M. 2018. The massive auditory lexical decision (mald) database. *Behavior Research Methods* 1–18.

[15] Vitevitch, M. S., Luce, P. A. 1998. When words compete: Levels of processing in perception of spoken words. *Psychological Science* 9(4), 325–329.

[16] Vitevitch, M. S., Luce, P. A. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40(3), 374–408.

[17] Vitevitch, M. S., Luce, P. A. 2004. A web-based interface to calculate phonotactic probability for words and nonwords in english. *Behavior Research Methods, Instruments, & Computers* 36(3), 481–487.

[18] Wood, S. 2017. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC 2 edition.