INVESTIGATING FREQUENCY OF OCCURRENCE EFFECTS IN L2 SPEAKERS: TALENT MATTERS

Jonathan Vais, Natalie Lewandowski, and Michael Walsh

Institute for Natural Language Processing, University of Stuttgart, Germany
firstname.lastname@ims.uni-stuttgart.de

ABSTRACT

This paper investigates the impact of word frequency and pronunciation talent on proficient speakers' second language (L2) production variability/similarity. The data analysed stem from a corpus of quasi-spontaneous conversations between native speakers of German and English. Within-speaker production similarity was established through a comparison of amplitude envelope signals for word tokens of L2 English word types. Production similarity scores, type frequency, and speaker talent were used as input for several linear (mixed effects) models. The models yielded a significant effect of type frequency, and a signficant interaction of type frequency and talent, for predicting similarity. The results are discussed in the context of exemplar-based category formation and indicate that less talented speakers are less capable of forming L2 categories.

Keywords: Exemplar Theory, language talent, word frequency

1. INTRODUCTION

Exemplar Theory, first introduced in psychology as a means to explore perception and categorisation [16, 20] has enjoyed considerable interest in recent years in the context of linguistics, e.g. with regard to speech production [21, 26] and perception [10, 14]. At its core is the idea that linguistic stimuli are stored as highly detailed episodes, e.g. rich in acoustic and contextual information, and are then employed in both speech production and perception. Thus, in perception, incoming percepts are categorised with respect to how similar they are to extant exemplars stored in exemplar clouds in memory, and, in production, stored exemplars act as production targets. The exemplar memory is constantly updated and is highly sensitive to effects of frequency and to recency.

With respect to frequency, frequent units are represented by many exemplars in memory and infrequent units by fewer. Effects due to frequency have been found across multiple linguistic domains [4, 5, 6, 7, 8, 15, 18, 20, 22, 27]. For example, in speech production, increased production practice of a given unit, implicitly captured by an increased number of exemplars, can lead to decreased production variability, known as entrenchment [21]. Thus, the frequency with which a linguistic item is produced by a speaker can be shown to affect how the item is realised. To date, however, exemplar-based studies which have uncovered such effects due to frequency have typically not included characteristics of the speakers involved. In this paper, therefore, we examine the effect of linguistic talent on speech production from an exemplar-based perspective. Specifically, we examine pronunciation talent in the context of English as a second language (L2).

Depending on an individual's phonetic aptitude, one can assume differences in the handling of incoming signals (input stage), as well as at the retrieval stage in speech production (e.g. [17]). Previous research looking at individual differences linked to aptitude has found that talented speakers were more agreeable, conscientious and empathy-ready [11] and high-aptitude speakers have also exhibited greater phonetic convergence [17]. One could also view talent from a more processing-oriented perspective, decomposing it into memory and phonemic coding abilities [25].

What has not been the focus of research on talent so far, is the question whether or how frequency of input, and its potential storage, matters in L2 production. Consequently, given that, to date, little if any research has explored the nature of language talent in the context of Exemplar Theory, in this paper we investigate the relationship between L2 pronunciation talent and exemplar frequency and their influence on word production variability/similarity.

In what follows, section 2 describes our L2 pronunciation database, section 3 presents a number of statistical models we employ to examine the relationship between exemplar frequency and L2 pronunciation talent, sections 4 discusses the results of these analyses and section 5 offers some concluding remarks and ideas for future work.

2. DATABASE

This paper employs phonetic convergence data discussed in [17]. The aim of the study was to examine convergence within conversations between native and nonnative speakers of English. Unlike phonetic convergence, which examines across-speaker production similarity, our study focuses on within-speaker similarity.

2.1. Subjects

For the purposes of the phonetic convergence study 20 native Germans (10 males, 10 females, 20-42 years of age), one General American speaker (male, 33 years of age) and one Standard Southern British English speaker (female, 57 years of age) were employed. Each of the 20 native Germans had an English dialogue with the American speaker and a dialogue with the British speaker. The recordings took place in a sound-attenuated room. All German participants learned English from the 5th grade of school onwards (from approx. 11 years of age). No participant spent longer than 3 weeks in an English-speaking country, nevertheless they were all proficient users of English. Following a suite of language talent tests [12, 13], e.g. delayed and direct pronunciation imitation, perception of prosody/accent, reading excercises using foreign accents, the participants were categorised into more and less phonetically talented L2 speakers, 10 participants in each category.

2.2. Elicitation of data

The data we employ below was elicited from more/less talented speakers, discussed above, in a Diapix task [9]. The participants of the dialog each got a picture. The pictures were similar, but there were some differences between them, which the participants had to find without seeing their partner's picture. Consequently, they were forced to solve this task through dialogue. The advantage of this approach is that there is a high interaction between the dialog partners.

2.3. Token representation and calculation of within-speaker pronunciation similarity

Each target word in the Diapix dialogues was transformed into an amplitude envelope representation. Firstly, the target word tokens were equated for root mean square amplitude and separated into four logarithmically spaced frequency bands. The amplitude envelopes for the four bands were then estimated using the Hilbert transform (see [26] for further details). By using amplitude envelopes the speech signals were not only compared at a static moment and for a specific feature, but encompassing all spectral information present evolving over time, adding a dynamic component to the analysis.

The similarity with which each speaker produced tokens – now represented as envelope signals – of the same type was then calculated, pairwise, using cross-correlation [1]. The estimated similarity of the two envelopes was determined for every band separately. The final 'match value', being the maximum of the previously established similarity estimates, ranges between 0 and 1 for each pair of tokens compared [17]. An average value was then calculated over all the token pair similarity scores for each word type.

2.4. Data sampling

Although the use of the Diapix approach yields dialogues that are quasi-spontaneous, they do, however, bias word type frequency to a considerable extent, due to the nature of the spot-the-difference task. Consequently, the frequencies of the word types are not representative of what one might expect in a normal dialogue. In addition, from an exemplar-theoretic perspective, they would not approximate the level of exemplar storage that one might expect. Therefore, the word type frequencies in the study below were substituted for their frequencies in the CELEX database [19]. It is important to note that the CELEX frequencies offer a better approximation of the frequencies and relative frequencies with which words are used, and consequently they can function as a rough proxy to the frequency of tokens one might expect to find in exemplar memory. Although CELEX is not based on L2 learner data, the participants in this study were proficient speakers of English and have had significant exposure to the language, and to the authors' knowledge no alternative approximation method of establishing exemplar frequencies exists.

In addition to the use of CELEX frequencies, only word types with 4 or more tokens were employed in the study in order to ensure a reasonable similarity estimate. Thus, the data set we employ contains 70 types and 288 tokens with CELEX frequencies ranging between 5 and 996, and each type has an associated similarity score, as discussd above.

3. MODELS

In order to investigate the relationship between frequency, pronunciation talent and production vari-

	Estimate	SE	t-value	Pr(>ltl)
(Intercept)	-0.096	0.002	-39.52	< 2e-16 ***
Frequency	0.007	0.002	2.838	0.0049 **

Table 1: Linear model results predicting production similarity as a function of word type frequency.

	Estimate	SE	t-value
(Intercept)	-0.096793	0.005002	-19.351
Frequency	0.007375	0.003329	2.216

Table 2: Linear mixed effects model (random effects incorporated) predicting production similarity as a function of word type frequency

ability/similarity a variety of linear and linear mixed models were applied to the dataset discussed above using functions provided in R [24, 2].

These models inform our understanding of this relationship by predicting the value of a dependent variable, in our case our similarity scores, as a linear combination of the fixed factors (e.g. frequency or talent) plus some intercept, plus random effects. Prior to fitting the models, the similarity scores and the frequencies were logged, z-scored, and centred. We checked for normality and homoskedasticity by visual inspections of plots of residuals against fitted values.

The first model is a simple linear model which examines the influence of *frequency* on *similarity* (Eq. 1). The linear model reported a significant p-value of $p < 0.01$. Table 1 presents the results. The positive estimate value for (log) frequency indicates that similarity increases (variability decreases) with increasing frequency.

(1) $\mathit{Similarity} \sim \mathit{Frequency}$

However, it is reasonable to anticipate certain random influences to be present depending on the individuals involved in the dialogues (speaker and partner) and the words being produced. Consequently, this baseline linear model was augmented. As random effects, the linear mixed effects model (Eq. 2) has intercepts for the *speaker*, *partner* and *word*. The results for this model are presented in Table 2.

(2) $$\mathit{Similarity} \sim \mathit{Frequency} + (1|\mathit{Speaker}) + (1|\mathit{Partner}) + (1|\mathit{Word})$$

To establish if the addition of random effects yielded a better fit to the data, the linear model and the linear mixed effects model were compared. Log likelihood ratio testing was not possible because the models were of different types (one simple linear, the other linear mixed effects). Instead we employed R's AIC function which produces Akaike's Information Criterion [1] scores for the two models. Unsurprisingly, the linear mixed effects model with random effects yielded a lower (better) AIC score than the simple linear model (-1044.515 < -1013.803).

The significant influence of frequency on variability indicates a process of entrenchment, in keeping with other exemplar-theoretic studies. In addition, however, we are also interested in the impact *talent* will have on *similarity*. The next model adds *talent* as an additional fixed effect (Eq. 3) to the model from Eq. 2. The results are presented in Table 3.

	Estimate	SE	t-value
(Intercept)	-0.094456	0.006588	-14.337
Frequency	0.007412	0.003336	2.222
Talent	-0.004573	0.008202	-0.558

Table 3: Linear mixed model (random effects incorporated) results predicting production similarity as a function of word type frequency and talent.

(3) $$\mathit{Similarity} \sim \mathit{Frequency} + \mathit{Talent} + (1|\mathit{Speaker}) + (1|\mathit{Partner}) + (1|\mathit{Word})$$

The t-value in Table 3, -0.558, would indicate no sigificant effect of talent on similarity. However, it is possible that talent interacts with frequency. Therefore in Eq. 4 we added an interaction of frequency and talent. The results are presented in Table 4.

	Estimate	SE	t-value
(Intercept)	-0.094213	0.006682	-14.100
Frequency	0.002784	0.003869	0.720
Talent	-0.004979	0.008219	-0.606
Frequency:Talent	0.011647	0.004357	2.673

Table 4: Linear mixed model (random effects incorporated) results predicting production similarity as a function of word type frequency, talent, and the interaction of the two.

(4) $$\mathit{Similarity} \sim \mathit{Frequency} + \mathit{Frequency} * \mathit{Talent} + (1|\mathit{Speaker}) + (1|\mathit{Partner}) + (1|\mathit{Word})$$

Inspection of the t-values in Table 4 indicates a significant interaction of frequency and talent. Likelihood ratio testing, comparing the models in equations 2 and 4, was possible in this case as both models are linear mixed effects models. A new model was considered better than its predecessor if the improvement was significant ($p < 0.05$) and if the AIC

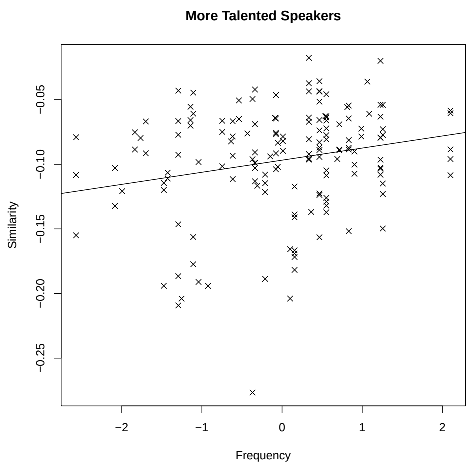
	Estimate	SE	t-value
(Intercept)	-0.097684	0.005214	-18.733
Frequency	0.010335	0.004205	2.458

Table 5: Linear mixed model (random effects incorporated) results predicting production similarity as a function of word type frequency, for talented speakers.

value was at least 2 points smaller [3]. The results demonstrated that inclusion of the interaction produced a significantly improved fit ($\chi^2 = 7.0417, p < 0.05$)

To investigate the nature of the signficant interaction between frequency and talent more closely, the data was divided into two sets consisting of the more talented (58 types, 158 tokens) and less talented speakers (52 types, 130 tokens), respectively. Subsequently, two linear mixed effects models, one for each talent data set were fitted according to Eq.2 to predict similarity as a function of frequency. The results for the more talented group are presented in Table 5. Once again, the positive (log) frequency estimate, 0.010335, indicates that with increasing frequency comes increasing similarity. The t-value in this case (2.458) indicates significance and a subsequent log likehood ratio test against a null model which only included the random effects demonstrated that the addition of frequency led to a significantly better fit of the data ($\chi^2 = 5.9494, p < 0.05$). The t-value for the less talented group (0.785) did not indicate significance and log likehood ratio testing did not yield a signficant improvement over the null model.

Figure 1: Word frequency plotted against similarity, for the more talented speakers.



4. DISCUSSION

In the context of previous exemplar-theoretic studies, the research presented above sought to establish whether word frequency impacted on production similarity/variability in L2 English and to investigate the role pronunciation talent might play. The results of the models indicate a significant entrenchment effect, whereby higher frequency of occurrence leads to lower variability [21]. In other words, production is refined with practice. Interestingly, this effect only holds for those participants who were categorised as talented with regard to pronunciation. Frequency does not appear to impact upon the realisation of tokens produced by the less talented grouping. Obviously, this invites the question as to why this might be.

Pierrehumbert notes that “beween physical experience and memory lies a process of attention, recognition, and coding which is not crudely reflective of frequency” [22, p.525]. Given that in our study frequency only appears to play a role for talented speakers, it might be that those speakers are better able to attend to appropriate L2 phonetic dimensions, to recognise L2 distinct exemplars (e.g. vowels not present in the L1 inventory) and to encode L2 exemplars in memory in exemplar clouds of their own. For the less talented speakers, lack of aptitude in these three areas would ultimately limit the extent to which robust L2 categories could emerge. Perceived L2 exemplars would be more likely to be categorised with respect to L1 exemplar clouds, and stored with their closest L1 equivalents. This would have the further consequence of L2 speakers being less capable of overcoming interference from their entrenched L1 exemplars. Consequently, one would not expect frequency to play a role in L2 production as the L2 exemplar clouds would be only sparsely populated, if present at all.

5. CONCLUSION

The research above demonstrates that talented L2 speakers of English are sensitive to the influence of word frequency on their production variability. This finding provides insights into the nature of L2 talent (e.g. the ability to attend to appropriate dimensions) and category formation. Future work will investigate frequency effects in other aspects of L2 proficiency likely to be affected by the dynamics of exemplar cloud formation.

6. ACKNOWLEDGMENTS

Funded under German Research Foundation (DFG) SFB 732

7. REFERENCES

[1] Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Petrov, B. N., Csaki, F., (eds), *2nd International Symposium on Information Theory* 267–281.

[2] Bates, D., Maechler, M., Bolker, B. 2013. *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999999-2.

[3] Burnham, K. P., Anderson, D. R. 2002. *Model selection and multimodel inference: a practical information -theoretic approach*. Springer 2 edition.

[4] Bybee, J. 10 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14, 261–290.

[5] Bybee, J. 2006. From usage to grammar: The minds response to repetition. *Language* 84, 529–551.

[6] Bybee, J., Scheibman, J. 1999. The effect of usage on degrees of constituency: the reduction of don't in English. *Linguistics* 37, 575–596.

[7] Carreiras, M., Perea, M. 2004. Naming pseudowords in spanish: Effects of syllable frequency. *Brain and Language* 90(1-3), 393 – 400. Third International Conference on the Mental Lexicon.

[8] Cholin, J., Levelt, W. J., Schiller, N. O. 2006. Effects of syllable frequency in speech production. *Cognition* 99(2), 205 – 235.

[9] van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., Bradlow, A. R. 2010. The wildcat corpus of native-and foreign-accented english: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech* 53(4), 510–540.

[10] Goldinger, S. D. 1997. Words and voices: Perception and production in an episodic lexicon. *Talker variability in speech processing* 33–66.

[11] Hu, X., Reiterer, M. S. 2009. Personality and pronunciation talent in second language acquisition. In: Dogil, G., Reiterer, S. M., (eds), *Language Talent and Brain Activity* Trends in Applied Linguistics. Berlin: de Gruyter 97–129.

[12] Jilka, M. 2009. Assessment of phonetic ability. In: Dogil, G., Reiterer, M. S., (eds), *Language Talent and Brain Activity*. Berlin: de Gruyter 17–66.

[13] Jilka, M. 2009. Talent and proficiency in language. In: Dogil, G., Reiterer, M. S., (eds), *Language Talent and Brain Activity*. Berlin: de Gruyter 1–16.

[14] Johnson, K. 1997. Speech perception without speaker normalization.

[15] Jurafsky, D., Bell, A., Gregory, M., Raymond, W. D. 2000. Probabilistic Relations between Words: Evidence from Reduction in Lexical Production.

[16] Kruschke, J. K. Jan. 1992. ALCOVE: an exemplar-based connectionist model of category learning. *Psychol Rev* 99(1), 22–44.

[17] Lewandowski, N. 2012. *Talent in nonnative phonetic convergence*. PhD thesis Universität Stuttgart.

[18] Losiewicz, B. L. 1992. *The Effect of Frequency on Linguistic Morphology*. PhD thesis University of Texas Austin, TX.

[19] Max Planck Institute for Psycholinguistics, 2001. Webcelex.

[20] Nosofsky, R. M. Mar. 1986. Attention, similarity, and the identification-categorization relationship. *J Exp Psychol Gen* 115(1), 39–61.

[21] Pierrehumbert, J. B. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In: Bybee, J. L., Hopper, P., (eds), *Frequency and the Emergence of Linguistic Structure*. John Benjamins Publishing Company 137–157.

[22] Pierrehumbert, J. B. 2006. The next toolkit. *Journal of Phonetics* 34(4), 516–530.

[23] Pluymaekers, M., Ernestus, M., Baayen, R. H. 2005. Lexical frequency and acoustic reduction in spoken dutch. *The Journal of the Acoustical Society of America* 118(4), 2561–2569.

[24] R Core Team, 2013. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.

[25] Skehan, P. 2003. *A cognitive approach to language learning*. Oxford applied linguistics. Oxford: Oxford Univ. Press.

[26] Wade, T., Dogil, G., Schütze, H., Walsh, M., Möbius, B. 2010. Syllable frequency effects in a context-sensitive segment production model. *Journal of Phonetics* 38(2), 227 – 239.

[27] Walsh, M., Möbius, B., Wade, T., Schütze, H. 2010. Multi-level exemplar theory. *Cognitive Science* 34, 537–582.

---

[1] Matlab's xcorr function was employed