Can the Discriminative Lexicon Model account for the family size effect in auditory word recognition?

Hanno Müller[1,2,*] (hanno.mueller@ru.nl), https://orcid.org/0009-0007-1784-5148

Louis ten Bosch[1] (louis.tenbosch@ru.nl), https://orcid.org/0000-0002-0152-9024

Mirjam Ernestus[1] (mirjam.ernestus@ru.nl), https://orcid.org/0000-0002-1853-0750

Affiliations:

[1]Centre for Language Studies, Radboud University, Nijmegen, Netherlands

[2]English Language and Linguistics, Heinrich Heine University, Düsseldorf, Germany

*Corresponding author

Abstract

Words with larger morphological families elicit shorter response times (RTs) in lexical decision experiments (e.g., Bertram et al. 2000). One possible account for this *family size (FS)* effect draws on the *Discriminative Lexicon Model* (DLM; Chuang & Baayen 2021), positing that morphological family members strengthen relationships between forms and meanings. While it has been shown that the DLM successfully explains FS effects in reading (Mulder et al. 2014), we investigated whether it does so in listening too. We trained the computational model LDL-AURIS (Shafaei-Bajestan et al. 2023), which implements the DLM, on Dutch and show that a measure derived from LDL-AURIS accounts for variance in auditory lexical decision RTs in Dutch, and also partially accounts for the same variance in the RTs as the auditory FS effect. Future research should investigate whether some other measure derived from the DLM can fully explain FS effects in listening.

Keywords: family size, discriminative lexicon model, auditory word recognition, morphological processing, lexical decision.

1. Introduction

A word's *family size* is the count of all word types in which the given word's root occurs as a constituent. For instance, family members of the word *think* are, among others, *doublethink, thinks, rethinking,* and *unthought*. In both visual and auditory lexical decision experiments, words with larger morphological families elicit quicker responses (e.g., Moscoso del Prado Martín et al. 2004; Müller et al. 2024). One explanatory account for this *family size effect* stems from the theory of the *Discriminative Lexicon Model* (DLM; Chuang & Baayen 2021). A computational implementation of the DLM successfully simulated the family size effect in the reading of English nouns (Mulder et al. 2014). It is not self-evident that the DLM can explain the family size effect in listening as well because the processes underlying written and spoken word recognition systematically differ. The present study investigates to what extent the DLM explains the auditory family size effect.

2. The family size effect

The family size effect has been observed in multiple studies and for multiple languages (e.g., Bertram et al. 2000; Mulder et al. 2014) in both reading and listening. It has been suggested that the family size effect is especially driven by semantic similarities between family members (e.g., Moscoso del Prado Martín et al. 2004). Moreover, family members contribute more to the effect the greater their similarity in form with the word to be recognised (Müller et al. 2024).

There are fundamental differences between how the family size effect manifests itself in visual and in auditory word recognition. In visual word recognition, the morphological structure of the word to be recognised does not affect the family size effect: the effect has been documented for prefixed (e.g., Moscoso del Prado Martín et al. 2004), simplex (e.g., Mulder et al. 2013), and suffixed words (e.g., Bertram et al. 2000). In contrast, in auditory word recognition, the family size effect is elicited only by simplex and suffixed words, but not by prefixed words, suggesting that the morphological structure of the word to be recognised interacts with the effect (Müller et al. 2024).

Differences between the visual and the auditory family size effect can be explained in light of systematic differences between reading and listening. During reading, a word's characters can all be simultaneously processed due to parafoveal preview (e.g., Rayner 1998).

This simultaneity renders it irrelevant whether the word's root, which is the shared element among the family members, occurs as first, second, or third constituent. In contrast, in auditory word recognition, words unfold over time and the word recognition process is assumed to start as soon as the audio input becomes available. Human auditory word recognition considers all words stored in memory that are (to some extent) compatible with the audio presented so far and gradually winnows out words that become more incompatible when the audio unfolds (e.g., Marslen-Wilson & Welsh 1978). Because of the incremental unfolding of the audio signal, prefixes are first perceived and processed, then roots, then suffixes. This may explain why the family size effect is less strong for prefixed words: the recognition process for prefixed words is already well under its way before the root, linking the members of a morphological family, becomes discernible.

3. Discriminative Lexicon Theory

Mulder and colleagues (2014) propose that the family size effect can be understood in terms of discriminative learning in the DLM. Discriminative learning supposes that the association between parts of a word's form and the word's meaning is strengthened when both occur together and weakened when one of them is present while the other is absent. The latter may occur, for instance, when a morpheme has several meanings or when a sound sequence (e.g., /rɛd/) is part of words with different meanings (e.g., *red*, *bread*). A stronger association leads to faster recognition when the word is presented. In the DLM, the family size effect is explained by the principle that the more family members a word has, the stronger the association between the word's root and its meaning.

The DLM is usually implemented as a two-layer neural network that takes as its input words' feature representations and predicts words' meaning representations. Feature representations can take the form of letter sequences or acoustic features, while meanings can be represented by arbitrary identifiers or semantic vectors (see below). For making predictions, the network first has to be trained, that is, it has to establish the association weights between word feature representations and meaning representations, on the basis of the input features and meanings of numerous words. When associations weights have been established, the DLM can determine a given input word’s meaning by comparing the support of the word’s features for all word meanings in the lexicon, modulated by the association strengths. Vice versa, the DLM can produce a word’s form, given a word’s meaning. Previous studies have shown that measures derived from the DLM can predict behavioural data including visual lexical decision data and acoustic durations (for an overview, see Chuang & Baayen 2021).

Mulder and colleagues (2014) implemented an early version of the DLM as a computational model for word reading, called the *Naïve Discriminative Reader*. The *Naïve Discriminative Reader* takes as its input character trigrams. Mulder and colleagues used the support for the meanings of the words in the lexicon to simulate lexical decision data. Family size was a significant predictor for the observed and the simulated lexical decision data, suggesting that the DLM can explain family size effects.

There is only one implementation of the DLM for spoken word recognition: LDL-AURIS (Shafaei-Bajestan et al. 2023). This implementation takes as its input words' audio recordings, of which the frequency spectra are summarised by means of *Continuous Frequency Band Summary Features* (C-FBSFs). Whereas the Naïve Discriminative Reader represents the meaning of a word with a unique letter sequence (localist representation), LDL-AURIS represents meanings with semantic vectors produced by a distributional semantics model. These vectors reflect that words with similar meanings tend to co-occur with the same set of words.

Shafaei-Bajestan and colleagues (2023) tested LDL-AURIS by determining how well it can recognise words sliced from continuous speech. This is a difficult task for human listeners, because, in everyday speech, words tend to be coarticulated and reduced, which makes them difficult to identify when presented without their contexts. Accordingly, human participants only identified 20.8% to 44.0% of the words sliced out of their contexts (Arnold et al. 2017). To determine whether LDL-AURIS correctly recognised a presented word, Shafaei-Bajestan and colleagues (2023) computed the correlation between the semantic vector computed for this word and all vectors in the lexicon. If the vector of the correct meaning was closest to the computed vector, the presented word was assumed to be correctly recognised. LDL-AURIS recognised 16% of the words, which approximated the lower bound of human performances in this difficult task. It has yet to be investigated whether LDL-

AURIS can also predict the time listeners need to recognise a word, for instance, the reaction times (RTs) from lexical decision experiments.

4. The present study

The present study focused on two research questions. First, we investigated whether LDL-AURIS can account for how quickly listeners recognise spoken words. More specifically, we investigated whether LDL-AURIS predicts RTs from an auditory lexical decision experiment. Second, we investigated whether and to what extent LDL-AURIS accounts for the same variance in the auditory lexical decision RTs as family size does.

Previous research has used different definitions of family size, resulting in different family size measures. In the present study, we focused on three of them. The first measure is *Classical Family Size*, for which all words including the presented word's root (i.e. family members) equally weigh. Like all family size measures, Classical Family Size yields a facilitative effect in lexical decision experiments (e.g., Bertram et al. 2000). The second measure is *Semantic Family Size*, for which the weight of a family member depends on the strength of its semantic relation with the presented word. Semantic Family Size yields better predictions for RTs than Classical Family Size (e.g., Moscoso del Prado Martín et al. 2004). Following Müller and colleagues (2024), the third measure is *Semantic Form Overlap Family Size*, for which the weight of a family member depends on both the strength of its semantic relation with the presented word and its form overlap with the presented word. Semantic Form Overlap Family Size is the best predictor of all family size measures for visual and auditory lexical decision RTs (Müller et al. 2024). For a detailed description of how we computed these family size measures, see Subsection 5.3.

We tested LDL-AURIS against the RTs from the Biggest Auditory Lexical Decision Experiment Yet (BALDEY; Ernestus & Cutler 2015), a Dutch large-scale auditory lexical decision experiment. We chose Dutch because Müller and colleagues (2024) showed that the three above-mentioned family size measures are statistically significant predictors of lexical decision RTs in Dutch.

In order to investigate our research questions, we compared three types of models, as summarised in Table 1. First, we built a statistical baseline model to predict the RTs from BALDEY that includes the most important control variables known to predict auditory lexical decision RTs (see Subsection 5.4), in order to decrease the variance in the RTs. We compared this baseline model with a model that also contained a predictor derived from LDL-AURIS (in interaction with the word's morphological structure). If the latter model is better, LDL-AURIS contributes to explaining the RTs.

Second, we produced three new statistical models by extending the baseline model with both the LDL-AURIS measure and a family size measure (both in interaction with morphological structure). We investigated whether any of these three new models better fit the RTs than an extension of the baseline model with just the LDL-AURIS measure. If so, the LDL-AURIS measure does not fully explain family size effects.

Third, we investigated whether the LDL-AURIS measure accounts for at least part of the family size effect. To this end, we investigated how much any of the three family size measures (in interaction with morphological structure) improves the model fit when added to the baseline model and compared this to how much any of the three family size measures improves the model fit when added to a model also containing the LDL-AURIS measure (in interaction with morphological structure). If the presence of the LDL-AURIS predictor results in a smaller improvement in terms of model fit, the LDL-AURIS measure accounts for at least part of the family size effect.

Table 1. Overview of model comparisons and the conclusions that can be drawn based on the results. SemDens refers to Semantic Density, the LDL-AURIS measure that we tested. Family Size represents any of the three tested family size measures.

Model 1	**Model 2**	**Interpretation of potential results**
Baseline	Baseline + SemDens	If Model 2 is better than Model 1, the LDL AURIS measure accounts for RTs.

Baseline + SemDens	Baseline + SemDens + Family Size	If Model 2 is better than Model 1, the LDL-AURIS measure does not (fully) account for family size effects.
Baseline (1a) vs. Baseline + Family Size (1b)	Baseline + SemDens (2a) vs. Baseline + SemDens + Family Size (2b)	If the difference in the model's goodness of fit with the RTs between Models 1a and 1b is larger than between Models 2a and 2b, the LDL-AURIS measure accounts at least to some extent for the family size effects.

Because LDL-AURIS has not yet been used to predict lexical decision RTs, we based the choice of our predictors on previous studies using the DLM to predict visual lexical decision RTs. Previous studies identified two predictors. The first predictor is *Target Correlation*, which is defined as the correlation between the semantic vector produced by LDL-AURIS on the basis of the audio input and the semantic vector of the correct word in the lexicon (Heitmeier et al. 2023a). Because this predictor did not predict RTs with statistical significance for our BALDEY dataset, we refrain from further discussing this predictor.

The second predictor is *Semantic Density*, which is the average cosine similarity between the semantic vector produced by LDL-AURIS based on the audio signal and each of the ten closest semantic vectors in the lexicon, in terms of cosine similarity. Heitmeier and colleagues (2023b) report that higher Semantic Densities correlate with shorter RTs. Their explanation for this finding is that when the semantic vector produced by the model lands in areas of more words, the presented word has a higher *wordlikeness*, which facilitates a "yes" response in lexical decision experiments.

5. Experiment

The data and the scripts that were used for this study can be downloaded from: https://doi.org/10.34973/x6v3-yj45.

5.1 Data

We predicted the RTs from BALDEY, which contains response latencies from 20 native speakers of Dutch to 2,780 spoken Dutch content words and 2,761 pseudowords. We only analysed correct responses to all real words, except for compounds, that were also part of the training set of LDL-AURIS (see Subsection 5.3). The dataset thus comprised 15,936 responses with 5,908 responses to 322 unique simplex words, 227 responses to 12 unique prefixed words, 8,875 responses to 478 unique suffixed words, and 926 responses to 50 unique words containing both a prefix and a suffix. We excluded 24 (0.15%) responses from Participant 1 in Session 8 due to an encoding error and the 248 responses (1.56%) given before stimulus offset.

5.2 Training LDL-AURIS

We trained LDL-AURIS on the audio recordings of *Component o* of the *Spoken Dutch Corpus* (Oostdijk 2000), which contains read-aloud speech from Dutch native speakers. We chose read-aloud speech because it is usually clearly pronounced, like the stimuli in BALDEY. Word tokens were sliced out from their acoustic context based on word segmentations as provided in the corpus. We removed mispronounced, incomplete, and unintelligible word tokens. The resulting dataset contains 550,688 word tokens (39,278 word types).

LDL-AURIS' input matrix specifies for each word token its acoustic properties in the form of C-FBSFs. The output matrix specifies for each word token its gold standard semantic vector, which is the semantic vector derived from a distributional semantics model. We used a Dutch distributional semantics model (Nieuwenhuijse 2018) that was trained on more than 600 million messages on Dutch social media, news, blogs, and forums, with word2vec (Mikolov et al. 2013). We removed 36,002 word tokens (14,646 word types) from the training data of LDL-AURIS, because the distributional semantics model did not provide semantic vectors for these words.

We trained LDL-AURIS in *julia* (Bezanson et al. 2017) with the package *JudiLing* (Luo et al. 2020). All parameters were exactly set as by Shafaei-Bajestan and colleagues (2023), who provide more details about the training procedure of LDL-AURIS.

5.3 Calculation of the family size measures

We determined all family size measures exactly as Müller and colleagues (2024). We based the family size measures on words incorporated in CELEX (Baayen and colleagues 1996).[1] Classical Family Size has a mean of 7.07 (SD = 0.36), Semantic Family Size has a mean of 7.19 (SD = 0.28), and Semantic Form Overlap Family Size's mean is 5.28 (SD = 0.98). For the statistical analyses, which were conducted with these measures, the measures were first log-transformed and then normalised with a z-transformation. After this preprocessing, Classical Family Size strongly correlates with Semantic Family Size (r = .998) and with Semantic Form Overlap Family Size (r = .804); Semantic Family Size and Semantic Form Overlap Family Size also strongly correlate (r = .810).

5.4 Control variables in the baseline model

Our baseline model included four control variables. The first is *Morphological Structure* (MorphStr) with the levels "prefixed", "simplex", "suffixed", and "double-affixed", because the auditory family size effect has been shown to vary with the morphological structure of the presented word (Müller et al. 2024). Second, the *Moving Average Response Time* (maRT) models the weighted average response time over preceding trials (ten Bosch et al. 2018). Third, to capture a participant's adjustment to the task throughout the entire experiment, we incorporated the number of each *Trial* (e.g., Ernestus & Cutler 2015). Fourth, we factored in form *Frequency*, which we obtained from CELEX (Freq; Baayen et al. 1996). All control variables were first log-transformed and then z-transformed for scaling and centring.

[1] Because LDL-AURIS was trained on fewer word types than are available in CELEX, we tested whether the results of this study change when the family size measures are only based on those that also occur in the LDL-AURIS training data. These alternative family size measures correlate with the family size measures reported in this paper with coefficients between .90 and .93 for any family size measure. More importantly, these alternative measures yield results very similar to those reported in this study. We chose to present the results from the family size measures based on the complete CELEX database in this study because we believe that they better reflect a listener's knowledge of words, which is not only based on listening but also on reading.

The correlation coefficients between each pair of control variable and between each control variable and Semantic Density or a family size measure is smaller than .1, except that Frequency weakly correlates with the three family size measures (r_{min} = .34, r_{max} = .38). These correlations are considered too low to be problematic.

5.5 Estimation and comparison of the models

We implemented all statistical models as *Generalised Additive Mixed Models*, with R 4.0.5 (R Core Team 2017) and the package mgcv (Wood 2015). We preferred this type of model over *Linear Mixed-Effects Models* (e.g., Bates et al. 2015), because the former can easily detect both linear and non-linear effects, whereas the latter can only detect linear effects. A practical introduction to Generalised Additive Mixed Models is provided by, for instance, Chuang and colleagues (2021).

We fitted all models in the style of Bates and colleagues (2015). That is, for the baseline model, the initial model comprised as predictors a) a parametric term for the categorical variable Morphological Structure, b) thin plate regression splines for all continuous control variables, c) a by-participant intercept, and d) by-participant random slopes for all continuous variables. We subsequently simplified this model by step-wise elimination of predictors that did not reach statistical significance. Then, we assessed whether pairs of predictors exhibited a concurvity exceeding 0.7, suggesting that half of the explained variance attributed to a given predictor is actually accounted for by other predictors. For pairs that surpassed the threshold of 0.7, we tested whether eliminating one of the two predictors impacted the other predictor's significance level (measured by the p-value) or shape (as depicted in an effects plot). If so, we eliminated the predictor with the smaller p-value. This procedure ensures that the model can accurately estimate all included predictors' effects (Tomaschek et al. 2018).

As summarised in Table 1, we compared the baseline model and models extended with Semantic Density or a family size measure. Following standard procedures (e.g., Chuang et al. 2021), we did so using a χ^2-test on likelihood scores for nested model comparisons. For comparing the difference in model improvement between two pairs of models, we compared decrease in AIC.

5.6 Results

Because the baseline model is not of interest by itself, we will not discuss it here. We just note that the control variables show approximately the same types of effects in the baseline model as in the best model developed in this study, which is summarised in Appendix A.

Adding Semantic Density in interaction with Morphological Structure as predictor to the baseline model results in a significantly better fit to the data ($\chi^2_{(2)} = 366.446$, $p < .001$). Semantic Density is a significant predictor of the RTs of simplex ($F = 66.237$, $p < .001$), suffixed ($F = 303.520$, $p < .001$), and prefixed words ($F = 4.891$, $p = .016$), but not of double-affixed words ($F = 0.591$, $p = .442$). This shows that a predictor derived from the DLM as implemented in LDL-AURIS can explain variance in auditory lexical decision RTs of words that are made of at most two morphemes. As illustrated in Figure 1, the effect of Semantic Density is inhibitory, that is, stimuli with a greater Semantic Density are responded to more slowly. The effect seems to level-off for values of Semantic Density greater than 0.6.[2]

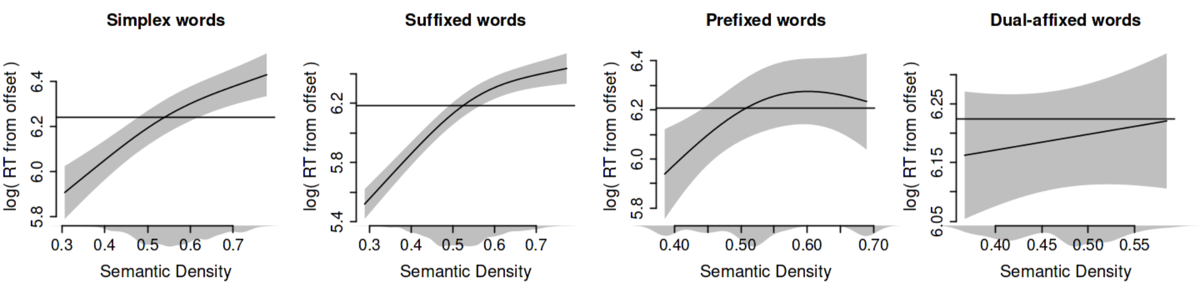

Figure 1. Partial effect of Semantic Density (x-axis) on log-transformed RTs (y-axis) for words with different morphological structures (panels). The density plots below the x-axes

[2] Based on a question from a reviewer, we tested whether including an interaction between semantic density and word frequency would lead to a significantly better model, which is not the case.

indicate the number of responses to words with the corresponding Semantic Density.

The further addition of an interaction between any of the three family size predictors by Morphological Structure to the model with Semantic Density results in an even better fit to the data (see Table 2), with Semantic Density again showing the same types of effects as illustrated in Figure 1. This shows that any family size measure explains variance in the RTs that is not explained by Semantic Density.

Table 2. Comparisons between a) the Baseline model plus Semantic Density by Morphological Structure and b) the Baseline Model plus Semantic Density by Morphological Structure and plus any family size predictor by Morphological Structure.

Family Size Predictor	**$X^2(8.00)$**	**p**
Classical Family Size	12.087	.002
Semantic Family Size	13.664	< .001
Semantic Form Overlap Family Size	28.680	< .001

A summary of the best model, which includes Semantic Form Overlap Family Size in interaction with Morphological Structure, is similar to the summaries of the other two models with family size predictors and can be seen in Appendix A. As shown in Figure 2, the effect of Semantic Form Overlap Family Size is facilitative, as expected, but the size of the effect varies with the word's morphological structure.

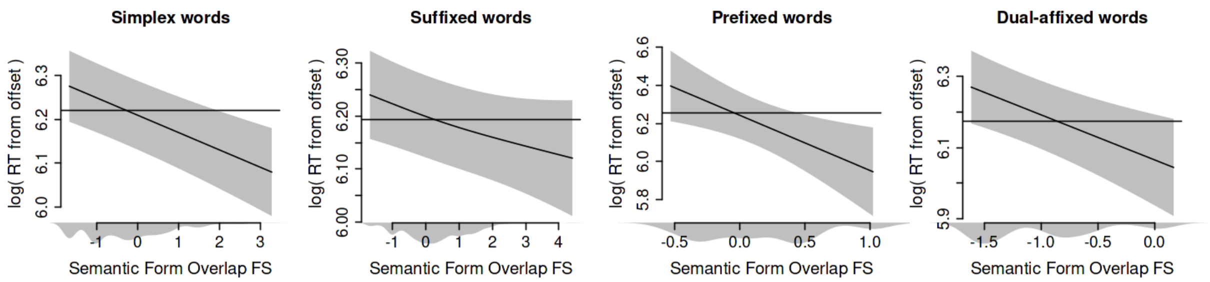

Figure 2. *Partial effect of log-transformed and centred Semantic Form Overlap Family Size (FS; x-axis) on log-transformed RTs (y-axis) for words with different morphological structures (panels). The density plots below the x-axes indicate the number of responses to words with the corresponding Semantic Form Overlap FS.*

Finally, Figure 3 shows that adding a family size predictor to the baseline model improves this model more than adding a family size predictor to a model that also contains the predictor Semantic Density. These results suggest that Semantic Density partially explains the same variance in the RTs as the family size predictors.

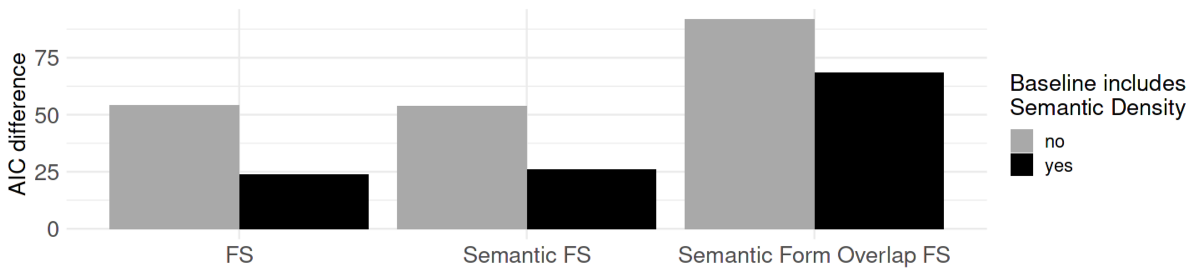

Figure 3. Improvement of the model's fit in terms of reduced AIC-points when either family size (FS) predictor is added to the baseline model (grey) and a model that also contains Semantic Density (black).

6. Discussion

This study has addressed the question whether the Discriminative Lexicon Model (DLM) can account for reaction times (RTs) from an auditory lexical decision experiment and for the family size effect in those RTs. We derived a measure, Semantic Density, from the computational model LDL-AURIS (Shafaei-Bajestan et al. 2023), which implements the DLM for auditory word recognition, and tested it against the RTs of the Dutch large-scale auditory lexical decision experiment BALDEY (Ernestus & Cutler 2015).

First, we investigated whether Semantic Density significantly accounts for variance in the RTs, which is the case. Our study thus enriches previous research that up to now has only shown that LDL-AURIS can recognise words approximately as accurately as human listeners (Shafaei-Bajestan et al. 2023), by showing that LDL-AURIS can also account for variance in auditory lexical decision RTs. Possibly, Semantic Density accounts for less variance in the RTs to words containing both a prefix and a suffix than to words of a simpler morphological structure due to sparseness of those complex words in LDL-AURIS' training set.

In our study, Semantic Density yielded an inhibitory effect. This may be surprising because Heitmeier and colleagues (2023b) reported a facilitative effect. One explanation for the inhibitory effect is already suggested by Heitmeier and colleagues: A higher Semantic Density implies that the meaning computed by LDL-AURIS is more similar to more words in the lexicon, which may render it more difficult to identify which of the meanings in the lexicon was intended. Our finding that a greater competition between meanings results in longer RTs is in line with cohort-driven auditory word recognition models such as DIANA (ten Bosch et al. 2022). The inhibitory effect in BALDEY suggests that in this experiment, participants only accepted a word as a real word when they knew the word’s meaning. In visual lexical decision experiments, participants may already have accepted a word because it was word-like, leading to a facilitative effect of Semantic Density.

Second, we investigated whether any of the three family size measures that we tested accounts for variance in the RTs that is not accounted for by Semantic Density, which is the case for all of them. Because this shows that Semantic Density does not fully account for the family size effect, we finally tested whether Semantic Density at least partially accounts for family size effects. For doing so, we tested whether adding any family size measure to a model containing Semantic Density improves the fit to the data less than adding this family size measure to a model without Semantic Density. This appeared to be the case, for all three family size measures. Therefore, our results suggest that Semantic Density at least partially accounts for the family size effects in listening. Our study therefore expands previous research by showing that not only the visual but also the auditory family size effect can at least partially be understood in terms of discriminative learning in the DLM.

LDL-AURIS relies on associations between forms and meanings. A given association is strengthened by more word tokens supporting this association (i.e., by more family members showing form overlap). Consequently, it may be expected that LDL-AURIS is most effective in explaining the effect of Semantic Form Overlap Family Size. As mentioned above, we tested for each family size measure to what extent it improves a model with and without Semantic Density. The more the addition improves a model without Semantic Density compared to the model with Semantic Density, the more effectively Semantic Density explains the effect of this family size measure. Contrary to expectations, Semantic Density explains the effect of Semantic Form Overlap Family Size to a lesser extent than the other two family size measures' effects. Apparently, the associations between form and meaning in the DLM contain slightly different information than Semantic Form Overlap Family Size. A probable cause is that associations between forms and meanings in the DLM are not only strengthened by word tokens supporting these associations, but also weakened by word tokens that do not support these associations, by representing similar forms but different meanings, or vice versa. Another probable cause is that LDL-AURIS is trained on word tokens whereas family size is based on word types. The association strength between a form and a meaning in LDL-AURIS can therefore represent different information from the Semantic Form Overlap Family Size of the word form.

Our study does not rule out that the auditory family size effect can be completely understood in terms discriminative learning in the DLM for auditory word recognition. Future research might derive a measure from the DLM that can account for the full variance explained by family size measures. Such a measure should probably not combine both positive and negative evidence for the association between forms and meanings in a single measure, like Semantic Density does, but purely reflect positive, morphological information.

In conclusion, our results show that the DLM contributes to explaining the variance in the RTs of an auditory lexical decision experiment. Moreover, the DLM can account for parts of the variance that is accounted for by family size measures. Future research has to show what this latter finding means for the DLM: whether a different measure can be derived from DLM implementations that can fully explain family size effects or whether the model first has to be adapted.

7. Funding information

This research was funded by the Deutsche Forschungsgemeinschaft (Research Unit FOR2373 ‘Spoken Morphology’, grant PL 151/7-2 ‘Central project’ to Ingo Plag, and grant ER 574/1-1 ‘Dutch morphologically complex words: The role of morphology in speech production and comprehension’ to Mirjam Ernestus, Louis ten Bosch, and Ingo Plag).

References

Arnold, Denis, Fabian Tomaschek, Konstantin Sering, Florence Lopez & Harald R. Baayen. 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. https://doi.org/10.1371/journal.pone.0174623

Baayen, R. Harald, Richard Piepenbrock & Leon Gulikers. 1996. The celex lexical database (cd-rom). *Linguistic Data Consortium*.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48. doi:10.18637/jss.v067.i01.

Bertram, Raymond, Harald R. Baayen & Robert Schreuder. 2000. Effects of family size for complex words. *Journal of Memory and Language, 42*, 390–405.

Bezanson, Jeff, Alan Edelman, Stefan Karpinski & Viral B. Shah. 2017. Julia: A fresh approach to numerical computing. *SIAM review, 59*(1), 65-98.

Chuang, Yu-Ying & Harald R. Baayen. 2021. Discriminative learning and the lexicon: NDL and LDL. In *Oxford research encyclopedia of linguistics*.

Chuang, Yu-Ying, Janice Fon, Ioannes Papakyritsis & Harald R. Baayen. 2021. Analyzing phonetic data with generalized additive mixed models. In *Manual of clinical phonetics*, 108-138.

Ernestus, Mirjam & Anne Cutler. 2015. Baldey: A database of auditory lexical decisions. *Quarterly Journal of Experimental Psychology, 68*, 1469–1488.

Heitmeier, Maria, Yu-Ying Chuang, Seth D. Axen & Harald R. Baayen. 2023a. Frequency effects in Linear Discriminative Learning. arXiv preprint arXiv:2306.11044.

Heitmeier, Maria, Yu-Ying Chuang & Harald R. Baayen. 2023b. How trial-to-trial learning shapes mappings in the mental lexicon: Modelling lexical decision with linear discriminative learning. *Cognitive Psychology*, *146*, 101598.

Luo, Xuefeng, Yu-Ying Chuang & Harald R. Baayen. 2020. JudiLing: an implementation in Julia of Linear Discriminative Learning algorithms for language model. Eberhard Karls Universität Tübingen, Seminar für Sprachwissenschaft.

Marslen-Wilson, William D. & Alan Welsh. 1978. Processing interactions and lexical access during word recognition in continuous speech. *Cognitive psychology, 10*(1), 29-63.

Mikolov, Tomas, Kai Chen, Greg Corrado & Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Moscoso del Prado Martín, Fermín, Raymond Bertram, Tuomo Häikiö, Robert Schreuder & Harald R. Baayen. 2004. Morphological family size in a morphologically rich language: The case of Finnish compared with Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1271.

Mulder, Kimberley, Ton Dijkstra, Robert Schreuder & Harald R. Baayen. 2014. Effects of primary and secondary morphological family size in monolingual and bilingual word processing. *Journal of Memory and Language*, *72*, 59–84.

Mulder, Kimberley, Robert Schreuder & Ton Dijkstra. 2013. Morphological family size effects in L1 and L2 processing: An electrophysiological study. *Language and Cognitive Processes, 28,* 1004–1035.

Müller, Hanno, Louis ten Bosch & Mirjam Ernestus. 2024. *The family size effect in visual and auditory word recognition. Language, Cognition and Neuroscience*, 1–22. https://doi.org/10.1080/23273798.2024.2337941

Nieuwenhuijse, Alexander. 2018. Dutch word2vec model. Retrieved 2022-02-10, from https://github.com/coosto/dutch-word-embeddings

Oostdijk, Nelleke 2000. The Spoken Dutch Corpus. Overview and First Evaluation. In *LREC* (pp. 887-894).

R Core Team. 2017. R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Rayner, Keith. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372.

Shafaei-Bajestan, Elnaz, Masoumeh Moradipour-Tari, Peter Uhrig & Harald R. Baayen. 2023. LDL-AURIS: A computational model, grounded in error-driven learning, for the comprehension of single spoken words. *Language, Cognition and Neuroscience*, 1-28.

ten Bosch, Louis, Mirjam Ernestus & Lou Boves. 2018. Analyzing Reaction Time Sequences from Human Participants in Auditory Experiments Proceedings of Interspeech 2018: The 19th Annual Conference of the International Speech Communication Association, pages 971-975.

ten Bosch, Louis, Lou Boves & Mirjam Ernestus. 2022. DIANA, a process-oriented model of human auditory word recognition. *Brain Sciences, 12*(5), 681-710.

Tomaschek, Fabian, Peter Hendrix & Harald R. Baayen. 2018. Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics, 71*, 249-267.

Wood, Simon. 2015. Package 'mgcv'. R package version, 1, 29.

Appendix A

Summary of the baseline model plus Semantic Density and Semantic Form Overlap Family Size and interactions, fitted to log-transformed RTs in our subset of BALDEY.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	6.22343	0.03595	173.127	< .001
MorphStrprefixed	0.03174	0.04600	0.690	.490
MorphStrsuffixed	-0.02794	0.01152	-2.837	.015
MorphStrdouble-affixed	-0.15014	0.05292	-2.424	.005
B. smooth terms	Edf	Ref.df	F-value	p-value
s(SemDensity:MorphStrsimplex)	1.855	1.979	75.543	< .001
s(SemDensity:MorphStrprefixed)	1.372	1.606	5.076	.006
s(SemDensity:MorphStrsuffixed)	1.978	1.999	286.310	< .001
s(SemDensity:MorphStrdouble-affixed)	1.000	1.000	0.006	.938
s(FamilySize:MorphStrsimplex)	1.000	1.000	20.716	< .001
s(FamilySize:MorphStrprefixed)	1.000	1.000	6.809	.009
s(FamilySize:MorphStrsuffixed)	8.116	8.784	6.005	< .001
s(FamilySize:MorphStrdouble-affixed)	1.000	1.000	12.308	< .001
s(Freq)	4.805	5.992	6.848	< .001
s(maRT)	4.869	6.084	302.558	< .001
s(Trial)	6.376	7.534	4.595	< .001
s(participant)	18.549	19.000	27.955	< .001
s(participant, Trial)	15.636	19.000	8.214	< .001

Corresponding address:

Hanno Müller, Postbus 9103, 6500 HD NIJMEGEN, THE NETHERLANDS