Annika Schebesta & Jessica Nieder

Contextual semantic effects in compound processing

Abstract: This study investigates how semantic transparency influences the phonetic realization of triconstituent nominal compounds (NNN) in English. Building on earlier work that employed averaged (static) embeddings, we use contextualized BERT representations derived from the exact carrier and context sentences presented to participants in a production experiment. This method aligns computational measures of semantic similarity with the experimental context guiding speakers' interpretations. Our analyses reveal a branching-sensitive asymmetry, indicating that contextualized semantic information affects phonetic realization selectively, depending on structural configuration. More broadly, the results highlight how computational embeddings can enrich experimental data by capturing semantic dimensions that are not directly accessible through traditional methods.

Keywords: compound processing, acoustic duration, distributional semantics, contextual embeddings

1 Introduction

Experimental linguistics has long relied on human behavioural and acoustic data to investigate how language is produced and processed. Following this tradition, phonetic and psycholinguistic studies typically explore language processing by analysing speech, response latencies, or judgments elicited in controlled experimental settings. Although this approach has successfully yielded rich insights into the structure of the mental lexicon, its reliance on observable behaviour often poses a challenge: crucial semantic information is frequently latent and must be inferred indirectly, or it needs to be incorporated into the study design in advance. Since many studies interested in morphological processing focus primarily on the interplay between morphology and phonetics (see e.g., Kunter and Plag, 2016; Plag et al., 2017; Schebesta and Kunter, 2022), an interplay that is already complex on its own, semantic factors are often not explicitly incorporated into experimental designs and, thus, remain unexamined.

Annika Schebesta & Jessica Nieder, Department of English linguistics, University of Siegen, Adolf-Reichwein-StraSSe 2, 57068 Siegen, e-mail: annika.schebesta@uni-siegen.de,

Recent advances in computational linguistics offer new tools to address this gap, e.g., in the form of post-hoc statistical analyses. This is particularly intriguing in those cases where semantic predictors were not included explicitly. In particular, distributional semantic models, for example those derived from large pre-trained neural language models such as Bidirectional Encoder Representations from Transformers (BERT, Devlin et al., 2019), provide vector-based representations that approximate aspects of human semantic knowledge. These computational embeddings have begun to enrich experimental analyses by quantifying semantic relationships that would otherwise remain opaque to traditional modelling (Buchanan et al., 2025; Nieder et al., 2024). As such, they offer a powerful way to operationalise theoretical constructs like semantic transparency in a quantifiable and replicable manner.

One linguistic domain where such enrichment of experimental data can be particularly valuable is compound processing (Schebesta and Nieder, 2024). Compounds, especially multi-constituent nominal compounds, pose unique challenges for theories of lexical representation. Their internal morphological structure often interacts with both phonetic realisation and semantic compositionality. This complexity is increased in triconstituent NNN compounds (e.g., *work accident factor*, *guest account service*), where multiple possible bracketing structures introduce ambiguity both in form and interpretation (Schebesta, 2024; Schebesta and Nieder, 2024). Crucially, these compounds additionally show varying degrees of semantic transparency, a factor long known for playing a pivotal role in compound processing (Creemers and Embick, 2022; Libben et al., 2003; Zwitserlood, 1994).

Previous research has shown that both the morphological structure and the phonological context affect the phonetic signal of compounds (Bell et al., 2021; Tomaschek et al., 2021; Schebesta and Kunter, 2022). More recently, Schebesta and Nieder (2024) demonstrated that semantic transparency, operationalised through word embeddings from a BERT model, predicts phonetic variation in triconstituent compounds. Specifically, they found that the acoustic duration of the third constituent (N3) in such compounds decreases as semantic transparency increases. These results provide novel computationally derived evidence that meaning, not just morphological or phonological factors, can influence phonetic realisation, thus complementing the earlier experimental findings on semantic transparency in compound processing.

However, the method used in the study by Schebesta and Nieder (2024) relied on static embeddings[1], which average over many contexts of use in the BERT model. This means that while the embeddings provide a general sense of meaning similarity, they fail to reflect the specific contextual semantics relevant to a given utterance. Human speakers, by contrast, interpret compounds in concrete discourse contexts that guide how constituent meanings combine and disambiguate. This is specifically important for experimental studies where speakers are often confronted with novel word forms or compound structures whose interpretation is intentionally shaped by context sentences designed to elicit specific morphological and semantic configurations such as different branching structures.

In the present study, we build on this prior work by moving from averaged or 'static' to contextualized BERT word embeddings. Using BERT, we extract vector representations of compound constituents as they appear in the exact experimental sentences seen by participants in the production study reported in Schebesta (2024) and Schebesta and Nieder (2024). This allows us to model semantic transparency in a way that more closely reflects the actual linguistic input that speakers processed during production. Crucially, it ensures that both human speakers and the computational model are exposed to the same contextual data, aligning their semantic representations more directly.

By incorporating these contextualised measures into a regression analysis of constituent durations, we investigate whether this more fine-grained semantic modelling enhances our understanding of the interaction between meaning and phonetic realisation. This approach offers a principled way to bridge experimental phonetics with distributional semantics and pushes forward the integration of computational models into psycholinguistic and phonetic research.

2 Experimental data

This study is based on experimentally elicited data from a production task with native speakers of North American English at the University of Alberta, Edmonton (Schebesta, 2024, ch. 8). The production study investigates whether and how the morphology-driven branching direction of English NNN compounds affects the phonetic signal of the three nominal compound constituents. That is, the order of compound formation is assumed to be determinative of the phonetic

1 Note that the embeddings were nevertheless derived from a contextual model. Schebesta and Nieder (2024) used the model in a 'static' capacity by retrieving the average embeddings directly from BERT.

realization of NNN compounds. The study tests the prediction that the morphologically embedded compound constituents are more phonetically reduced than the pertinent free compound constituent.

Previous research has shown that the branching direction of compounds containing three nominal constituents (N1, N2, and N3) is co-determined by the lexical bigram frequency of compound constituents and the overall morphological structure of the compounds (Kunter and Plag, 2016; Schebesta and Kunter, 2022). This is reflected in statistically significant three-way interactions between the constituent number, the morphological branching and the lexical bigram frequency. In order to disentangle this effect, the NNN compounds in the production task were designed in such a way that the lexical bigram frequencies of the bigrams N1N2 and N2N3 were very low, i.e., < 12 hits in the reference corpus COCA (Davies, 2008). Doing so, an effect of lexical frequency on the phonetic signal of morphologically complex words (see e.g., Hay, 2001; Bien et al., 2005; Hay, 2007) is ruled out to the largest possible extent. A third determinant of branching direction, namely the orthographic form of a compound (Sanchez-Stockhammer, 2018), was controlled for by spelling all bigrams involved in the experimental study with spaces.

In total, 25 W1W2 pairs such as *accident*$_{\mathrm{W1}}$ *factor*$_{\mathrm{W2}}$ were formed and served as the basis of NNN compounds. Each W1W2 pair was accompanied by a third - preceding or following - nominal constituent, resulting in 50 different NNN compounds, e.g., *work*$_{\mathrm{N1}}$ *accident*$_{\mathrm{W1}}$ *factor*$_{\mathrm{W2}}$ and *accident*$_{\mathrm{W1}}$ *factor*$_{\mathrm{W2}}$ *survey*$_{\mathrm{N3}}$.

The newly created NNN compounds (which did not exist in COCA) were embedded in short text passages. These contained one context sentence, whose semantics determined the branching direction of the NNN compounds, and a carrier sentence. (1) illustrates the text passage for the left-branching NNN compound [*work*$_{\mathrm{N1}}$ *accident*$_{\mathrm{N2}}$] *factor*$_{\mathrm{N3}}$: In the context sentence, the embedded constituents *work*$_{\mathrm{N1}}$ and *accident*$_{\mathrm{N2}}$ form part of the sentence-initial Noun Phrase, while the free constituent *factor*$_{\mathrm{N3}}$ follows at the end. The distances between the three compound constituents is supposed to suggest that the NNN compound is left-branching. The context sentence was followed by the carrier sentence that holds the NNN compound *work*$_{\mathrm{N1}}$ *accident*$_{\mathrm{N2}}$ *factor*$_{\mathrm{N3}}$ in Subject position.

(1) The *accidents* at *work* are in particular related to one *factor*.
The [*work*$_{\mathrm{N1}}$ *accident*$_{\mathrm{N2}}$] *factor*$_{\mathrm{N3}}$ has been analysed by an expert.

For each of the 50 NNN compounds, two text passages were created in such a way that each NNN compound occurred once as left-branching and once as right-branching in the data set. The 42 participants (37 female, 5 male) were recorded which reading aloud the 100 text passages containing the compounds.

After a manual acoustic analysis of the speech recordings with PRAAT (Boersma and Weenink, 2017), the data was statistically analysed using R (R Core Team, 2023). A linear mixed-effects regression model (Bates et al., 2015) predicted constituent durations and segment durations. The statistical model included a number of phonological (pitch measurements, number of phonemes and syllables), lexical (unigram frequency, phonological neighborhood density) and extra-linguistic (speech rate, repetitions of NNN) noise variables, and incorporated two random intercepts for the speaker and the compound constituent. The analysis showed that in both left- and right-branching NNN compounds, N1 and N2 constituents have the same duration and are significantly shorter than N3 constituents. The findings suggest that the morphological structure of NNN compounds does not manifest itself in the phonetic signal, whereas most of the noise variables show the predicted effect: More accented constituents are longer than unaccented constituents, constituents are produced shorter after three and four repetitions, and constituents with a high number of phonological neighbors are produced shorter than those with fewer phonological neighbors. The question arises whether semantics, which has been shown to be indicative of phonetic variation, can be used to explain the duration pattern N1, N2 $<$ N3 in NNN compounds.

3 Method

3.1 Computational pre-processing

To obtain contextualised measures of semantic transparency, we represented each constituent by the final-layer `[CLS]` embedding of a masked carrier+context input, treating this sentence-level vector as a proxy for the constituents contextualised meaning. Following Schebesta and Nieder (2024) and Buijtelaar and Pezzelle (2023), we employed the pre-trained `bert-base-uncased` model from the `Hugging Face Transformers library` to generate vector representations for each target word. The embedding pipeline was implemented in Python using `PyTorch`. Embeddings were processed using `NumPy` and stored in `.csv` format. These vectors were later used to compute cosine similarity-based semantic transparency.

Each NNN compound in the data set, e.g., *work accident factor*, was accompanied by a context sentence and a carrier sentence from the original experimental design. These were concatenated into a single string that served as the input to the BERT model. The complete BERT input for the left-branching interpretation of the compound *work accident factor* was thus:

(2) **BERT input**: The accidents at work are in particular related to one factor. The *work accident factor* has been analysed by an expert.

This approach ensured that the computational model accessed the same information as human participants, maintaining a consistent representational basis. For each compound, we extracted embeddings for its three constituents. For the left-branching compound *work accident factor*, this resulted in a word embedding vector for the free constituent *factor* ($\mathbf{v}_{\text{free}}$), a vector for the first embedded constituent *work* ($\mathbf{v}_{embed_1}$) and a vector for the second embedded constituent *accident* ($\mathbf{v}_{embed_2}$).

To isolate each constituents contextualized meaning, we created three masked versions of the full input by replacing one constituent at a time with a `[MASK]` token. For example, to extract the contextual representation of *factor*, the input string was modified to:

(3) The accidents at work are in particular related to one factor. The *work accident* `[MASK]` has been analysed by an expert.

Each masked input was passed through BERT, and the output embedding corresponding to the `[CLS]` token in the final hidden layer was extracted as a summary representation of the sentence with the masked word. This process was repeated for all three constituents in every item. The resulting 768-dimensional CLS embeddings for each constituent were stored alongside metadata (constituent labels, context and carrier text, masking success) until all embeddings were successfully extracted.

3.2 Semantic transparency calculation

To quantify semantic transparency (ST), we followed the cosine similarity-based approach introduced by Buijtelaar and Pezzelle (2023) adapted to triconstituent compounds by Schebesta and Nieder (2024).

We first computed an average embedding for the two embedded constituents ($\mathbf{v}_{embed_1}$ and $\mathbf{v}_{embed_2}$), and then calculated the cosine similarity between this combined vector and the embedding of the free constituent. This value represents how well the meaning of the free constituent aligns with the semantic content of the embedded pair:

(4)

$$\mathrm{ST}(c) = \frac{\mathbf{v}_{\text{free}} \cdot \left(\frac{\mathbf{v}_{embed_1} + \mathbf{v}_{embed_2}}{2}\right)}{\|\mathbf{v}_{\text{free}}\| \cdot \left\|\frac{\mathbf{v}_{embed_1} + \mathbf{v}_{embed_2}}{2}\right\|}$$

In addition to the composite ST score, we also computed pairwise cosine similarities between each pair of constituents:

(5)

$$\mathbf{v}_{\text{embed}_1} - \mathbf{v}_{\text{free}}$$
$$\mathbf{v}_{\text{embed}_2} - \mathbf{v}_{\text{free}}$$
$$\mathbf{v}_{\text{embed}_1} - \mathbf{v}_{\text{embed}_2}$$

All cosine similarity computations were performed with L2-normalized vectors. In the final dataset, all items contained valid embeddings for all three constituents. The similarity measures were added to the dataset for subsequent statistical modelling.

3.3 Statistical modelling

We ran a linear mixed-effects regression model in R (R Core Team, 2023) using the R package `lme4` (Bates et al., 2015). The model in (6) illustrates the final model, which predicts the Duration of the three constituents N1, N2 and N3 in seconds (sec). The random intercepts Speaker and Constituent account for speaker-specific effects and individual variation introduced by a given constituent into the regression model. In order to address our research question whether semantic similarity impacts the constituent duration of left- and right-branching NNN compounds, it is inevitable to incorporate the predictor Member containing the three levels `N1`, `N2` and `N3`, which correspond to the three compound constituents, and the predictor rated Branching, which has the levels `left-branching` and `right-branching`. rated Branching indicates the branching direction of each of the 100 NNN compounds as it has been assessed in a rating task with 38 native speakers of North American English (Schebesta et al., prep). Thirdly, the semantic similarity of compound constituents is encoded in three different similarity measures which, however, strongly correlated during preliminary modeling. In order to minimize the risk of collinearity, we opted for a Principal Component Analysis (James et al., 2013), which generated the output variables PC1 cosineSimilarity and PC2 cosineSimilarity. Thus, the statistical model incorporates the two three-way interactions Member × PC1 cosineSimilarity × rated Branching and Member × PC2 cosineSimilarity × rated Branching. With these interactions, the statistical model is able to predict a potential effect of the semantic similarity between embedded and free constituents on N1, N2 and N3, even if the effect is different for `left-branching` and `right-branching` NNN compounds.

(6) Duration ∼ (1 | Speaker) + (1 | Constituent)
+ Member × PC1 cosineSimilarity × rated Branching
+ Member × PC2 cosineSimilarity × rated Branching
+ lgUnigramFreq
+ Speech Rate
+ Repetition
+ PC1 phonological setup + PC2 phonological setup
+ PC1NNN phonological segments
+ PC1 pitch + PC2 pitch + PC3 pitch

The statistical model is complemented with a group of noise variables, namely lgUnigramFreq which corresponds to the lexical frequency of individual constituents (log-transformed to the base 10), the Speech rate at which a NNN compound was produced, and Repetition which indicates the number of repetitions of the contained W1W2 pairs (max. four repetitions per W1W2 pair, max. two repetitions per NNN). In addition, a set of pitch measurements, that have been generated during the acoustic analysis, were incorporated in the regression model in order to account for a potential effect of prominence patterns or accentuation on the duration of constituents. Information about the phonological neighborhood density of N1, N2 and N3 was generated from the *English Lexicon Project* (Balota et al., 2007) and encoded in two additional predictor variables. Another four predictors informed about the number of phonological segments (i.e. phonemes and syllables) of both the whole NNN compound as well as the compound constituents. However, due to strong correlations between (i) the pitch measurements, (ii) the two predictors with the number of phonological segments in NNN compounds, and (iii) the predictors informing about the phonological length and the number of phonological neighbors of the individual compound constituents, we performed three Principal Component Analyses (James et al., 2013) in order to avoid collinearity. Thus, instead of the input variables, the Principal Components PC1 phonological setup and PC2 phonological setup inform about the phonological setup of the compound constituents, PC1NNN phonological segments indicates the number of phonological segments of the whole NNN compounds, and PC1 PITCH, PC2 PITCH and PC3 PITCH contain information about the pitch contours of compound constituents.

The distribution of residuals of the initial model did not meet the normality assumption of linear regression so that the dependent variable DURATION was Box-Cox-transformed (Box and Cox, 1964) using an exponent of $\lambda = 0.02$. 65 outliers from a total of 10,710 observations from 3,573 NNN compounds were excluded from the data set (0.6%). After the reduction of the data set and the transformation of the dependent variable DURATION, we used the `vif()` function from the `car` package (Fox and Weisberg, 2019) to check for correlated predictors, which did not reveal any collinearity. In order to minimize the risk of Type I errors and overfitting, we did not reduce the regression model further (see Harrell, 2001; Winter, 2020 for a discussion). Table 1 provides a summary of the dependent and independent variables incorporated in the initial regression model, including the Principal Components.

4 Results

The results of our linear mixed-effects regression model ($R^2_{\text{marg.}}$ 0.566, $R^2_{\text{cond.}}$ 0.802, AIC -96550.363) are displayed in Table 2. The coefficient estimates indicated in the first column correspond to the predicted Box-Cox-transformed duration of N1, N2 and N3 constituents. The Intercept indicates the estimated constituent duration for the underlying reference level MEMBER=`N1`, RATED BRANCHING=`left`, REPETITION=`1`.

Fig. 1 illustrates the statistically significant partial effects of the noise variables SPEECH RATE, REPETITION, PC1 PHONOLOGICAL SETUP and PC2 PITCH on constituent duration on the y-axis (in seconds). The significant effects of SPEECH RATE and REPETITION go in the expected direction: As can be seen in the top row of fig. 1, a higher speaking rate of the participants and more repetitions of a W1W2 pair lead to shorter constituent durations. Note that with each repetition cycle, the duration of compound constituents decreases significantly compared to the previous repetition. The results are in agreement with previous findings on shortening effects of SPEECH RATE (Johnson et al., 1993) and REPETITION (Jacobs et al., 2015; Kahn and Arnold, 2015).

The bottom-left plot displays a lengthening effect of PC1 PHONOLOGICAL SETUP on constituent durations. According to the loadings in that Principal Component, an increase of that variable indicates an increase of the number of phonemes and syllables and the simultaneous decrease of the number of phonological neighbors of a given compound constituent. Thus, the effect of PC1

Dependent variable	Min, Max	Median	Mean	SD
Duration (transformed)	0.96, 1.00	0.98	0.99	0.01
Predictor variables - numeric				
lgUnigramFreq	2.95, 5.61	4.48	4.50	0.46
Speech Rate	3.04, 11.22	6.79	6.81	1.58
PC1 phonological setup	-4.40, 5.33	0.11	0.01	1.76
PC2 phonological setup	-3.39, 1.01	0.15	0.00	0.75
PC1NNN phonological segments	-2.18, 3.17	0.04	0.00	1.38
PC1 pitch	-5.20, 6.97	-0.37	0.79	1.87
PC2 pitch	-5.60, 2.85	0.36	0.00	1.21
PC3 pitch	-5.61, 4.96	-0.02	0.00	0.96
PC1 cosineSimilarity	-0.94, 9.54	-0.27	0.00	1.41
PC2 cosineSimilarity	-2.57, 4.36	-0.23	0.00	0.98
Predictor variables - categorical		Levels	N	Percentage
Repetition		1	2,691	25.3%
		2	2,668	25.1%
		3	2,589	24.3%
		4	2,697	25.3%
rated Branching		left	8,832	71.2%
		right	1,813	28.8%
Member		N1	3,564	33.0%
		N2	3,541	33.7%
		N3	3,540	33.3%
Interactions				
Member × PC1 cosineSimilarity × rated Branching				
Member × PC2 cosineSimilarity × rated Branching				
Random effects				
Speaker				
Constituent				

Tab. 1: Summary of variables contained in the initial regression model predicting NNN constituent durations (N=10,645).

Variables of interest	Estimate	Std. Error	*t*	*p*	
(Intercept)	0.986	0.002	493.625	<0.001	***
Member=N2	0.000	0.000	-4.403	0.000	***
Member=N3	0.002	0.000	15.326	<0.001	***
PC1 cosineSimilarity	0.000	0.000	-0.567	0.571	
PC2 cosineSimilarity	0.000	0.000	-0.073	0.942	
rated Branching=right	0.000	0.000	-1.733	0.083	
rated Br=right:Member=N2	0.001	0.000	2.784	0.005	**
rated Br=right:Member=N3	0.000	0.000	1.346	0.178	
PC1 cosSim:rated Br=right	-0.001	0.001	-1.474	0.141	
PC1 cosSim:Member=N2	0.000	0.000	0.254	0.799	
PC1 cosSim:Member=N3	0.000	0.000	-2.497	0.013	*
PC1 cosSim:rated Br=right:Member=N2	0.002	0.001	1.984	0.047	*
PC1 cosSim:rated Br=right:Member=N3	0.002	0.001	1.953	0.051	
PC2 cosSim:rated Br=right	0.000	0.000	1.117	0.264	
PC2 cosSim:Member=N2	0.000	0.000	-0.796	0.426	
PC2 cosSim:Member=N3	0.000	0.000	0.266	0.790	
PC2 cosSim:rated Br=right:Member=N2	0.000	0.001	-0.683	0.495	
PC2 cosSim:rated Br=right:Member=N3	0.000	0.001	-0.141	0.888	
Noise variables					
lgUnigramFreq	-0.001	0.000	-1.323	0.189	
Speech Rate	0.000	0.000	-3.709	<0.001	***
Repetition=2	0.000	0.000	-3.829	<0.001	***
Repetition=3	-0.001	0.000	-9.406	<0.001	***
Repetition=4	-0.001	0.000	-12.357	<0.001	***
PC1 phonological setup	0.002	0.000	17.917	<0.001	***
PC2 phonological setup	0.000	0.000	-1.037	0.303	
PC1NNN phonological segments	0.000	0.000	0.280	0.779	
PC1 pitch	0.000	0.000	0.044	0.965	
PC2 pitch	0.000	0.000	-9.382	<0.001	***
PC3 pitch	0.000	0.000	0.169	0.866	

Tab. 2: Fixed effects in final regression model (dependent variable: Box-Cox transformed Duration, N=10,645, reference level of the Intercept: rated Branching = left, Member = N1, Repetition = 1).

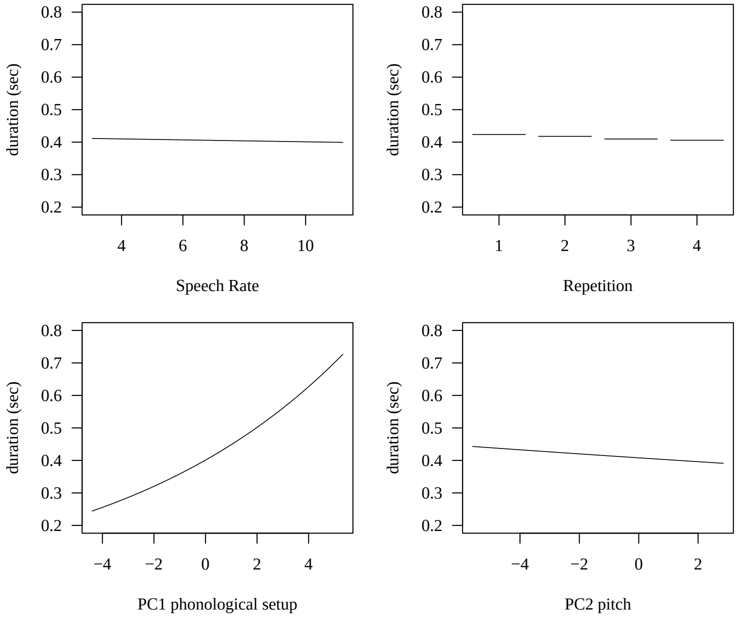

Fig. 1: Partial effects of SPEECH RATE, REPETITION, PC1 PHONOLOGICAL SETUP, and PC2 PITCH on DURATION (back-transformed (in sec), numeric predictors at median, categorical predictors at most frequent category).

PHONOLOGICAL SETUP is in accordance with previous research which reports more phonetic reduction in lexical units in denser phonological neighborhoods (Gahl et al., 2012).

In the bottom-right plot, the effect of PC2 PITCH on constituent duration is illustrated. Note that the Principal Component is mostly informed by a negative loading of accentuation, that is, the decrease of PC2 PITCH corresponds to an increase of accentuation of a given compound constituent. Accordingly, accented constituents with a lower PC2 PITCH value are overall longer than unaccented constituents with a higher PC2 PITCH value, which has also been reported in Turk and White (1999).

The effect of semantic similarity (on the x-axis) on the DURATION of compound constituents (on the y-axis) in the two branching directions is displayed in Fig. 2. Note that PC1 COSINESIMILARITY can be interpreted in such a way that a low value (left edge in each plot) indicates more similarity between the compound constituents, and a high value (right edge in each plot) indicates less similarity. The top row illustrates the interaction plots for the three left-branching compound constituents N1, N2 and N3, and the right-branching constituents N1, N2 and N3 are presented in the bottom row. Although a first glimpse at the regression lines might suggest otherwise, there is no statistically significant difference between the left-branching constituents and the pertinent right-branching counterparts. That is, the effect of PC1 COSINESIMILARITY is as strong on N1 in left-branching as in right-branching compounds, and the same holds for N2 and N3.

Within the left-branching NNN compounds, N3 constituents are significantly longer when the compound constituents are semantically more similar. In contrast, N1 and N2 are unresponsive towards semantic similarity. Apparently, the free constituent behaves differently from the embedded constituents in left-branching NNN compounds. In right-branching compounds, semantic similarity has no significant effect on the duration of any of the three compound constituents. Although the slope of all three regression lines appears to be reasonably steep, the right-branching compound constituents are unresponsive towards PC1 COSINESIMILARITY. As a result, the behavior of the right-branching embedded and free constituents differs from that of the left-branching compounds reported before.

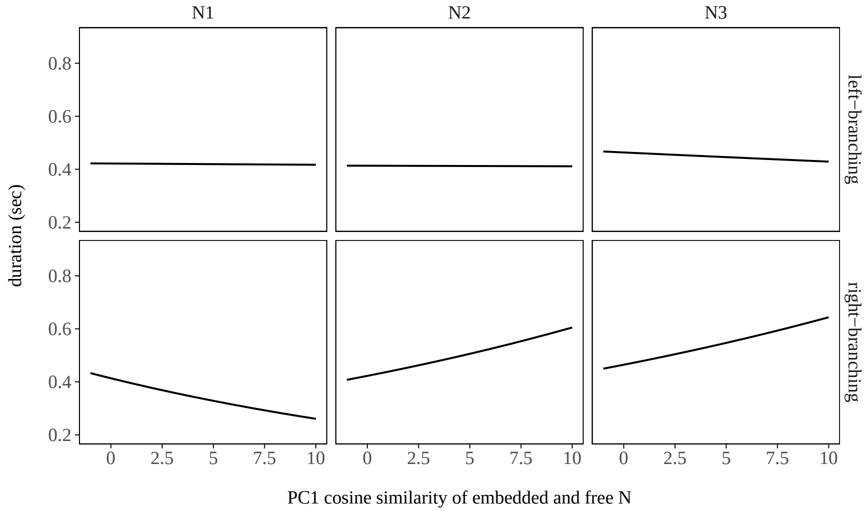

Fig. 2: Effect of MEMBER × PC1 COSINESIMILARITY × RATED BRANCHING on DURATION (back-transformed (in sec), numeric predictors at median, categorical predictors at most frequent category).

5 Discussion

The present study set out to examine whether contextualized semantic embeddings can account for phonetic variation in triconstituent NNN compounds more precisely than averaged embeddings. Compared to earlier work (Schebesta and Nieder, 2024), which relied on averaged (or 'static') BERT representations, the present results suggest a more differentiated role of semantics in compound processing.

In the study by Schebesta and Nieder (2024), semantic transparency was found to predict reduced durations of the free constituent (N3): the more transparent a compound, the shorter N3 became. This pattern pointed to a global effect of meaning similarity, independent of branching direction. By contrast, our contextualized approach revealed that semantic influences are more selective. We observed a significant effect only in left-branching compounds, where N3 lengthened as the embedded and free constituents became more semantically similar. Right-branching compounds, in turn, showed no semantic effect at all.

This branching-sensitive asymmetry highlights two important points. First, it indicates that contextualized embeddings, which mirror the actual input available to participants, do not merely replicate static similarity measures but instead capture semantic factors that are shaped by the syntactic-morphological configuration. Second, it suggests that the free constituent in left-branching compounds is particularly sensitive to semantic integration, potentially because the embedded pair forms a more cohesive semantic unit that places interpretive or processing demands on N3. The absence of a corresponding effect in right-branching compounds may reflect the looser semantic cohesion of the initial constituents in that configuration.

Crucially, this finding aligns with the observation that left-branching structures were not only preferred by participants in the rating task, but are also the preferred pattern in English compounding more generally. Indeed, previous research has shown that left-branching NNN compounds are generally more frequent than right-branching NNN compounds in English: Huber (2023) shows in her usage-based investigation of English NNN compounds that 63% of the compounds in her data set are left-branching and 37% are right-branching NNN. According to the semantic analysis in Kösling and Plag (2009), 73% of the NNN compounds in their data set (elicited from the Boston University Radio Speech Corpus, (Ostendorf et al., 1997) are left-branching while 27% are analysed as right-branching. The stronger sensitivity of left-branching compounds to semantic similarity may therefore reflect the fact that these structures constitute the default configuration in English, and thus provide a more stable basis for speakers to integrate semantic information into their phonetic realization.

Taken together, these findings refine our understanding of how meaning modulates the phonetic signal in complex words. While both of our studies converge on the idea that semantic relationships cannot be ignored in compound processing, the present results emphasize that such effects are conditional on structural context. More broadly, they demonstrate the utility of contextualized embeddings as a tool for linking distributional semantics with experimental data.

6 Limitations

Our approach necessarily operationalizes semantic relationships in a specific way, namely through the `[CLS]` representation of masked carrier+context sentences. This choice provides a consistent, usage-based summary of the constituents meaning, but it also means that our measures reflect one particular projection of distributional information in the BERT model. Other ways of representing constituent meaning (e.g., directly at the token position without masking) may highlight different aspects of semantic relatedness. More generally, distributional embeddings are proxies for meaning: they can capture certain dimensions of semantic similarity but cannot exhaustively represent the interpretive processes available to speakers. The present findings should therefore be understood as evidence that contextualized semantic cues can play a role in shaping the phonetic realization of compounds, without implying that the specific embedding method employed here captures all linguistically relevant aspects of transparency. We leave it up to future studies to explore other embedding methods.

Bibliography

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., and Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39(3):445–459.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Bell, M. J., Ben Hedia, S., and Plag, I. (2021). How morphological structure affects phonetic realisation in English compound nouns. *Morphology*, 31:87–120.

Bien, H., Levelt, W. J. M., and Baayen, R. H. (2005). Frequency effects in compound production. *Proceedings of the National Academy of Sciences of the United States of America*, 102(49):17876–17881.

Boersma, P. and Weenink, D. (2017). Praat: Doing phonetics by computer. http://www.praat.org/.

Box, G. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, 26(2):211–256.

Buchanan, E. M., Cuccolo, K., Heyman, T., van Berkel, N., Coles, N. A., Iyer, A., et al. (2025). Measuring the Semantic Priming Effect Across Many Languages. https://doi.org/10.31219/osf.io/q4fjy. OSF Preprint.

Buijtelaar, L. and Pezzelle, S. (2023). A Psycholinguistic Analysis of BERT's Representations of Compounds. In Vlachos, A. and Augenstein, I., editors, *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 2230–2241, Dubrovnik, Croatia. Association for Computational Linguistics.

Creemers, A. and Embick, D. (2022). The role of semantic transparency in the processing of spoken compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 48(5):734–751.

Davies, M. (2008). The Corpus of Contemporary American English (COCA). https://www.english-corpora.org/coca/.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Burstein, J., Doran, C., and Solorio, T., editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Fox, J. and Weisberg, S. (2019). *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, third edition.

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

Harrell, F. E. (2001). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. Springer eBook Collection Mathematics and Statistics. Springer, New York, NY.

Hay, J. (2001). Lexical frequency in morphology: Is everything relative? *Linguistics*, 39(6):1041–1070.

Hay, J. (2007). The phonetics of 'un'. In Munat, J., editor, *Lexical creativity, texts and contexts*, volume 58 of *Studies in Functional and Structural Linguistics*, pages 39–57. Benjamins, Amsterdam.

Huber, E. (2023). *Tri-Constituent Compounds. A Usage-Based Account of Complex Nominal Compounding*. De Gruyter Mouton, Berlin, Boston.

Jacobs, C. L., Yiu, L. K., Watson, D. G., and Dell, G. S. (2015). Why are repeated words produced with reduced durations? Evidence from inner speech and homophone production. *Journal of Memory and Language*, 84:37–48.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning with applications in R*. Springer, New York, 6th printing 2015 edition.

Johnson, K., Ladefoged, P., and Lindau, M. (1993). Individual differences in vowel production. *The Journal of the Acoustical Society of America*, 94(2 Pt 1):701–714.

Kahn, J. M. and Arnold, J. E. (2015). Articulatory and lexical repetition effects on durational reduction: speaker experience vs. common ground. *Language, Cognition and Neuroscience*, 30(1-2):103–119.

Kösling, K. and Plag, I. (2009). Does branching direction determine prominence assignment? An empirical investigation of triconstituent compounds in English. *Corpus Linguistics and Linguistic Theory*, 5(2):201–239.

Kunter, G. and Plag, I. (2016). Morphological embedding and phonetic reduction: the case of triconstituent compounds. *Morphology*, 26:201–227.

Libben, G., Gibson, M., Yoon, Y. B., and Sandra, D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language*, 84(1):50–64.

Nieder, J., van de Vijver, R., and Ussishkin, A. (2024). Emerging Roots: Investigating Early Access to Meaning in Maltese Auditory Word Recognition. *Cognitive Science*, 48:e70004.

Ostendorf, M., Price, P., and Shattuck-Hufnagel, S. (1997). *Boston University Radio Speech Corpus LDC96S36*. Linguistic Data Consortium, Philadelphia, PA.

Plag, I., Homann, J., and Kunter, G. (2017). Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics*, 53(1):181216.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Sanchez-Stockhammer, C. (2018). *English compounds and their spelling*. Studies in English language. Cambridge University Press, Cambridge.

Schebesta, A. (2024). *NNN compounds in English: Investigating the interface of morphology, lexical frequency and the phonetic signal*. PhD thesis, Universität Siegen. [Unpublished doctoral dissertation].

Schebesta, A. and Kunter, G. (2022). Constituent durations in English NNN compounds: A case of strategic speaker behavior? *Journal of Phonetics*, 94.

Schebesta, A. and Nieder, J. (2024). Semantic transparency affects the phonetic signal. https://doi.org/10.31234/osf.io/xtm9j. OSF Preprint.

Schebesta, A., Nieder, J., and Saito, M. (in prep.). Predicting compound branching directions with distributional semantics.

Tomaschek, F., Plag, I., Ernestus, M., and Baayen, R. H. (2021). Phonetic effects of morphology and context: Modeling the duration of word-final S in English with naïve discriminative learning. *Journal of Linguistics*, 57(1):123–161.

Turk, A. E. and White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics*, 27(2):171–206.

Winter, B. (2020). *Statistics for linguists: An introduction using R*. Routledge Taylor & Francis Group, New York and London.

Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes*, 9(3):341–368.