Title

Non-syntactic Processing Explains Cortical Entrainment During SpeechPerception

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Author

Frank, Stefan L.

Publication Date

2017

Peer reviewed

Non-syntactic Processing Explains Cortical Entrainment During Speech Perception

Stefan L. Frank (s.frank@let.ru.nl)
Centre for Language Studies, Radboud University
P.O. Box 9103, 6500 HD Nijmegen, The Netherlands

Jin-Biao Yang (ray306@gmail.com)
Institute of Brain and Cognitive Science, NYU Shanghai
3663 North Zhongshan Road, Shanghai, China 200062

Keywords: Language processing; cortical entrainment; distributional semantics

Introduction

There is considerable debate on the precise role of hierarchical syntactic structure during the comprehension of sentences, with some arguing that a full hierarchical analysis is required for comprehension (e.g., Ding, Melloni, Tian, & Poeppel, 2017) and others claiming that non-hierarchical processing is more common (e.g., Frank, Bod, & Christiansen, 2012). Ding, Melloni, Zhang, Tian, and Poeppel (2016) recently presented evidence that cortical entrainment during speech perception reflects the neural tracking of hierarchical syntactic structure of simple sentences, which would support the view that hierarchical processing is unavoidable. However, we show that the same entrainment effects appear in a computational model that does not incorporate syntax or any other linguistic knowledge or process beyond the word level. Hence, the cortical entrainment results do not need to be indicative of syntactic processing.

Ding et al. (2016) had participants listen to Chinese or English four-syllable sentences, with syllables presented at a fixed rate while cortical activity was recorded with MEG. A frequency analysis of the MEG signal revealed peaks in the power spectrum at exactly the occurrence frequencies of syllables, phrases, and sentences. For example, when sentences with [NP VP] structure (such as “dry fur rubs skin”) were presented at a rate of 250 ms per monosyllabic word, peaks would appear at 4 Hz, 2 Hz, and 1 Hz, corresponding to the syllable/word, phrase, and sentence rate, respectively (see Figure 1). Likewise, a sequence consisting of only NPs or only VPs resulted in peaks at 4 Hz and 2 Hz but not 1 Hz, while presenting a sequence of Chinese syllables without word or phrase structure resulted in only the 4Hz peak.

Although these results can indeed be interpreted in term of the sentences’ syntactic structures, we propose a simpler explanation: The power spectrum merely reflects responses to regularities in word-level properties, such as (approximate) syntactic or semantic category. For example, in the [NP VP] sentences, verbs occur at 1Hz and nouns at 2Hz. We implemented this alternative explanation in a simple computational model and show that it indeed predicts the MEG power spectra in different experimental conditions.

The model

The only linguistic knowledge available to the model is encoded in word vector representations. These were generated by a distributional semantics model (Mikolov, Chen, Corrado, & Dean, 2013) trained on large corpora of Chinese or English texts (the same model and English corpus were used to obtain word representations that Frank & Willems, in press, and Frank, 2017, applied to account for N400 and reading time effects). Words that occur in similar contexts get similar vectors so that representations of words from the same syntactic/semantic category tend to be clustered together.

The stimuli from the (Ding et al., 2016) experiments were presented to the model at a simulated rate of 4 Hz per English word or Chinese syllable. Twelve different subjects were simulated by retraining the distributional semantic model and randomly varying stimuli presentation order. The sequence of vector representations, at a simulated time resolution of 5 ms, were analysed by applying a Discrete Fourier Transform to obtain a power spectrum, just like Ding et al. (2016) do in their analysis of the MEG signal.

Results

Figure 2 shows that the model predicts the same peaks in the power spectrum as in the original MEG study. The minor peak at 3 Hz, which did not reach significance in the MEG data, is most likely merely the second subharmonic of the 1 Hz peak (Zhou, Melloni, Poeppel, & Ding, 2016). The model further correctly accounts for the outcomes of experiments with two-syllable NP or VP sequences that lack full sentence structure, and predicts results very similar to those in the MEG data when syllable sequences are scrambled to remove any higher linguistic structure (see Frank & Yang, 2017)

Conclusion

The only linguistic knowledge in the model is encoded in the input vectors, so it remains at the lexical level. Furthermore, the model does not include any intergrative processing. The resulting power spectra can therefore not reflect any (hierarchical) syntactic processing. Consequently, the original MEG results may also be explained without syntax.

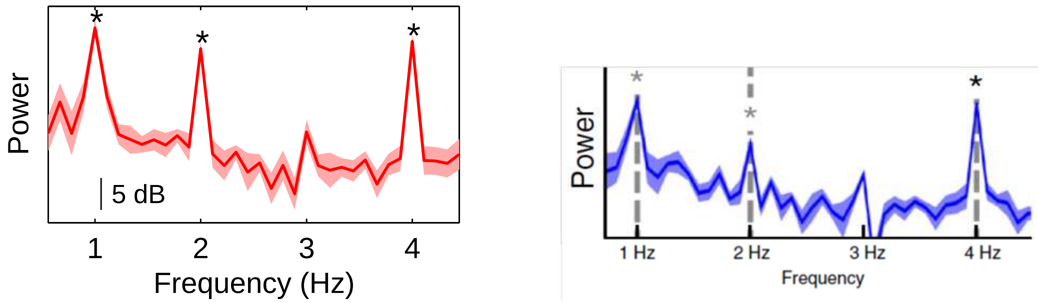

Figure 1: Left: MEG results for Chinese [NP VP] sentences. Right: MEG results for English [NP VP] sentences, reproduced from Ding et al. (2016, Figure 2e) with permission (the frequency scale was adapted to match simulated presentation rate). Shaded areas represents the standard error over subjects; lines are the average over subjects. Stars indicate significant peaks after multiple comparison correction.

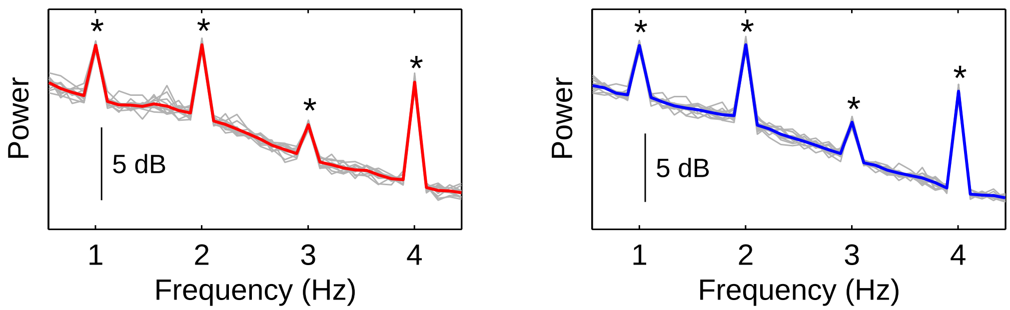

Figure 2: Model results for Chinese (left) and English (right) [NP VP] sentences. Grey lines represent individual simulated subjects; coloured lines are the averages over simulated subjects. Stars indicate significant peaks after multiple comparison correction.

References

Ding, N., Melloni, L., Tian, X., & Poeppel, D. (2017). Rule-based and word-level statistics-based processing of language: insights from neuroscience. *Language, Cognition and Neuroscience*, *32*, 570-575.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*, 158–164.

Frank, S. L. (2017). Word embedding distance does not predict word reading time. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society.*

Frank, S. L., Bod, R., & Christiansen, M. H. (2012). How hierarchical is language use? *Proceedings of the Royal Society B: Biological Sciences*, *279*, 4522–4531.

Frank, S. L., & Willems, R. M. (in press). Word predictability and semantic similarity show distinct patterns of brain activity in language comprehension. *Language, Cognition and Neuroscience*.

Frank, S. L., & Yang, J. (2017). Lexical representation explains cortical entrainment during speech perception. *Manuscript submitted for publications*.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In *Proceedings of the ICLR Workshop.*

Zhou, H., Melloni, L., Poeppel, D., & Ding, N. (2016). Interpretations of frequency domain analyses of neural entrainment: periodicity, fundamental frequency, and harmonics. *Frontiers in Human Neuroscience*, *10*, 274.