Differences between L1 and L2 processing of meaningful sentences revealed by combining EEG and eye tracking

Mulder, K.; Brand, Sophie; Boves, Lou

Publication date
2023

Document Version
Final published version

Published in
Interspeech

DIFFERENCES BETWEEN L1 AND L2 PROCESSING OF MEANINGFUL SENTENCES REVEALED BY COMBINING EEG AND EYE TRACKING

Kimberley Mulder, Sophie Brand, Lou Boves

Amsterdam Center for Language and Communication, University of Amsterdam; Radboud Teachers Academy; Center for Language Studies, Radboud University Nijmegen
k.mulder2@uva.nl; Sophie.Brand@ru.nl; Lou.Boves@ru.nl

ABSTRACT

We investigate how native French listeners and advanced Dutch learners of French use visual and auditory information while processing meaningful spoken sentences containing a reduced noun in mid-sentence position. Using a version of the visual world paradigm combined with EEG recordings, we investigate whether these reduced nouns affect the processing –as evidenced by the EEG signals– of a semantically related noun at the end of the sentence. We also investigate whether a newly developed feature that summarizes the eye movements in some time window affects the processing of those words. It appears that the reduction status of the first noun affects the processing of the related word and that this effect is different for the two listener groups. We also find that adding the new eye movement feature explains a highly significant amount of the variance in the EEG signals.

Keywords: visual world paradigm; EEG; eye tracking; reduced pronunciation; L1-L2

1. INTRODUCTION

Understanding reduced words in everyday speech can pose difficulties for second language learners, even if they are highly proficient in that language [1, 2], but little is known about the cognitive processes that make understanding reduced words more difficult for learners than for native listeners. To investigate this, we designed a visual world setting aimed to uncover those processes and their time course during listening comprehension, by combining EEG recordings and eye tracking. Native French listeners and advanced Dutch learners of French listened to meaningful sentences that contained a reduced or full form noun towards the middle of the sentence and a noun that was semantically related to that target noun towards the end of the sentence. Listeners who struggle to recognize a reduced target noun might need to use other sources of information to eventually correctly understand the message. This information could be semantically related information that occurs at the end of the sentence and the visual information about the target provided by the pictures. Through the simultaneous recordings of EEG and eye tracking we hope to see how native and non-native listeners deal with reduction over time and how they use visual information and semantically related words to solve any processing problems caused by reduction. We expect that effects of reduction last for a longer period of time for learners than for natives. In addition, the visual information (the pictures that appear on the screen one second before the start of the sentences and remain there until sentence end) that could help interpreting the reduced words might be used differently by natives and learners. If the reduced forms are not recognized and not matched with a visual target, this could affect they way in which semantically related word and the sentence as a whole are processed.

In order to explore long-distance effects, we need to go beyond standard analysis based on EEG time-locks on the target noun in the sentence. In [3] EEG signals were time-locked to fixations during the presentation of the audio signals and the presentation of the pictures, in addition to the start of the target noun, where the effect of reduction (perhaps better: the interaction between reduction and L1/L2) was assumed to be largest. In this study we focus on saccade data, and we investigate whether eye movements in the preview interval and in intervals aligned with the target and the related word affect ERP responses.

2. METHOD

We analyze the eye tracks and EEG signals obtained from 27 L2 and 28 native French listeners, who processed 115 stimulus sentences, the first three of which were practice trials. All sentences had the same structure: subject + verb form (+ que 'that') + target noun (+ verb form) + prepositional phrase / noun (e.g. *Elle a pris la tenaille pour enlever le clou*, "She took the pliers to remove the nail.") The subject and verb form at the start of the sentences were chosen to minimize the predictability of the target word that could be pronounced in full, or with a heavily reduced ə in the first syllable. From 1000 ms before the start of the spoken sentences up to their offset, the listeners saw four pictures on a screen, one of which corresponded to the target word. After the end of a sentence the four pictures were replaced by a photo of a scene, and the listeners had to decide whether that scene corresponded to the message conveyed by the sentence.

The 54 target nouns were all concrete nouns that can easily be depicted and comprise two or three syllables, the first of which contains a ə. The 58 filler nouns in 'target position' were one-, two- or three-syllable words that can easily be depicted. The semantically related nouns at sentence end were not subject to phonetic criteria. All stimuli were produced by a female speaker born and raised in the Northern part of France. Due to the phonetic and visual restrictions imposed on the target nouns it was not possible to impose additional restrictions, such as frequency of occurrence. The `lemFreqFilm` counts of the target nouns in the Lexique database [4] vary between 0.1 and 739.12, $\mu = 50.14$, $\sigma = 127.45$, and 0.34 -1061,92, $\mu = 113.56$ and $\sigma = 239.24$, for the related nouns.

We simultaneously recorded EEG (using an Acticap with 64 active electrodes and BrainVision© hardware/software) and eye movements (using an SR Research Ltd. Eyelink 1000 eye tracker). To ensure temporal synchronization of the EEG and eye signals within Presentation (www.neurobs.com), an eye-tracking extension was developed in-house.

The distance to the screen was such that the four pictures could easily be identified without the need for fixations. Actually, we found that fewer than half of the trials contained a fixation of at least 100 ms [5]) on any picture, both during the preview period and during a time interval of 1000 ms aligned with the onset of the target noun. Therefore, we developed a novel feature that summarizes five measures that can be derived from the eye movement: (1) the number of pictures fixated, (2) the total time the eyes are in a Region of Interest of any picture, (3) the number of saccades, (4) the total distance travelled by the eyes, (5) the area of the screen covered by the eyes, defined as the area of the ellipse with radii equal to the standard deviation of the X- and Y-coordinates. It appeared that we could compress the five measures using Principle Component Analysis. The first component alone explains more than 95% of the variance in the data.

The EEG epochs were subjected to automatic Independent Component Analysis (ICA) [6] to remove artifacts. Epochs that contained too many artifacts were discarded.

To remove the impact of exogenous auditory excitation and eye movements we subtracted the Multivariate Temporal Response Function (mTRF) [7] from the EEG signals. For that purpose we constructed an artificial excitation function that is the sum of the Hilbert envelope of the audio signal and the 1-D eye movement function representing the distance travelled by the eye gaze.

We limit the analysis to the 35 sensors that are assumed to be least affected by eye movements. Furthermore, we average the signals from individual sensors in nine subsets that cover the scalp in left-center-right in one and front-center-back in the other direction.

The impact of factorial and numerical predictors on the EEG signals for the factors of interest, here the reduction status of the target word and the listener group (L1 or L2), or perhaps better: on the interaction between the factors `reduction` and `language`, changes during the course of a stimulus [8]. However, we are interested in the time course of possible effects of control factors such as the frequency of the target and the related words, as well as of the predictors derived from the eye track signals. For that reason, we apply Linear Mixed Effect Models, using the `lme4` [9] package in `R` version 4.1.2 [10] to overlapping 100 ms windows shifting along the time axis, cf. [11, 12]. Here, we use 20 such windows; the first one starts at the onset of the target or related word; the last one ends at 1000 ms after word onset. Using partially overlapping time windows may give rise to the repeated analysis problem. To counteract that problem we only consider an effect as truly significant if its t-value is significant in at least three consecutive time windows.

To investigate the impact of predictors derived from the eye tracks we start with a basic model:

```
model.0 = lmer(amplitude~Alday+trial+
block+marker*language*logFreqFilm+time+
(1+logFreqFilm|ppn)+(1|word),data=window,
```

in which `amplitude` represents the 10 EEG samples in the window; `Alday` is the average amplitude of the EEG signals in a 240 ms interval preceding the onset of the word that replaces the conventional baseline correction [13]; `trial` is the number of the trial in either `block` #1 or #2 in a session; `marker` denotes the reduction status of a stimulus (full or reduced); `language` denotes the listener group; `logfreqFilm` is the log of the frequency count of the target word (when analyzing the data with time-lock on the onset of the target word) or of the related word (when time-lock is on the related word) in *Lexique* and `time` corresponds to the sequence of the samples in a window. `ppn` is the code of the 27+28 participants, and `word` the lexical identity of the target word. Factor levels `reducedFull` and `languageDutch` are on the intercept.

We then add additional predictors derived from the eye movement data, and use window-based AIC values to determine these improve the model fit, in which case we consider them as relevant.

3. RESULTS

We analyzed the features derived from the eye tracks to see whether there are significant differences between the two listener groups and/or between the three windows (preview, target, or related word) using t-tests for independent samples. In all three time intervals the distance traveled by the eyes of the L2 listeners is much larger than for the L1 listeners, but these movements appear to be constrained in a much smaller area of the display. Also, the L1 listeners fixate more often on a picture during the target window, and in the two other windows their fixations are longer. Thus, it appears that the two listener groups use the visual information in quite different manners.

Fig. 1 summarizes one detail of the analysis of the EEG signals related to the target word (red symbols) and the related word (blue symbols) in 20 time windows shown along the horizontal axis, viz. the t-values of the three-way interaction `reduction:language:frequency`. The nine panels show the results for the nine scalp areas, from left to right and front (top) to back (bottom). Note that the horizontal axis is labeled 'time from time-lock': the absolute time differs between target and related word. The vertical axes represent the t-values of the interaction `reduction:language:frequency`. The brown horizontal dashed lines indicate the value $t = 1.96$, and the yellow dotted line the value $t = -1.96$. The red '+' symbols show the windows in which the interaction is considered as significant in a brain area for the target words; the blue '*' symbols show the same data for the semantically related words.

Note that the factor `frequency` is tied to the word under analysis. Therefore, the predictors for the target and semantically related words are different. The results show a significantly more positive P200 component for the words with higher frequency in the native listeners, both for the reduced target

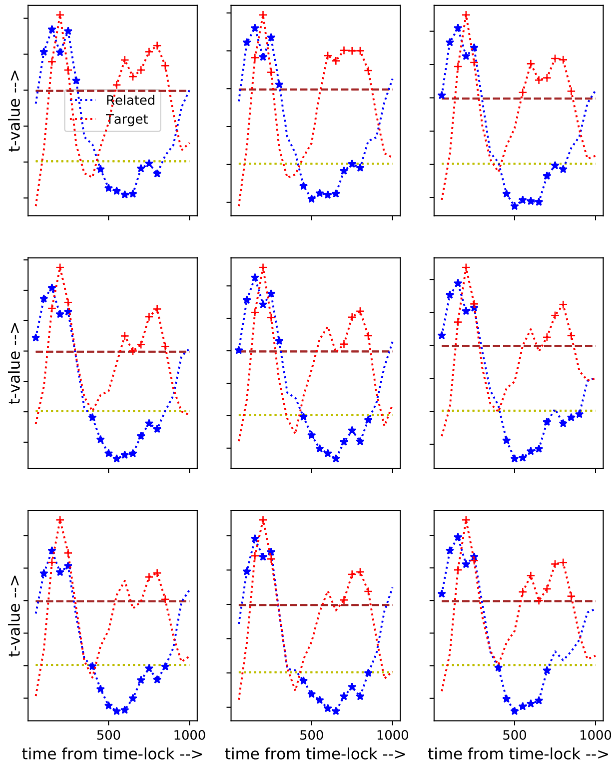

Figure 1: Three-way interaction `reduction:language:frequency` for the target and for the related word.

nouns, and -surprisingly- also for the semantically related words. This happens in all brain areas. For the target nouns there is also a significantly more positive late and somewhat broad P600 effect. The related words at the end of the sentences show a very different pattern. Here, in a broad time window starting 400 ms after word onset and lasting until 700 ms after onset, high frequency words following a reduced target show a significantly negative-going ERP.

Next, we built LME4 models for the target and the related nouns in which we added the eye track features as additional predictors. For the target nouns we built models with the eye track feature obtained in the 1000 ms preview period and aligned with the onset of the noun; we also added both predictors simultaneously. The latter model yielded by far the smallest AIC values in all 20 window positions. For the related noun we also added the eye track feature of the window aligned with that word, plus all combinations of the three features. Here too, the model that included all three eye track features had by far the smallest AIC in all 20 window positions.

Given the substantial contributions of the eye track features to the fit of the `lme4` models one would expect that adding these features should affect the significance of the factors that we are interested in, i.e., `reduction`, `language` and `frequency`. However, that appeared not to be the case. It appears that eye movement predictors that we add to the models do account for substantial amounts of variation in the EEG signals, but it is almost exclusively variance that has little or nothing to do with the linguistic factors that we are interested in.

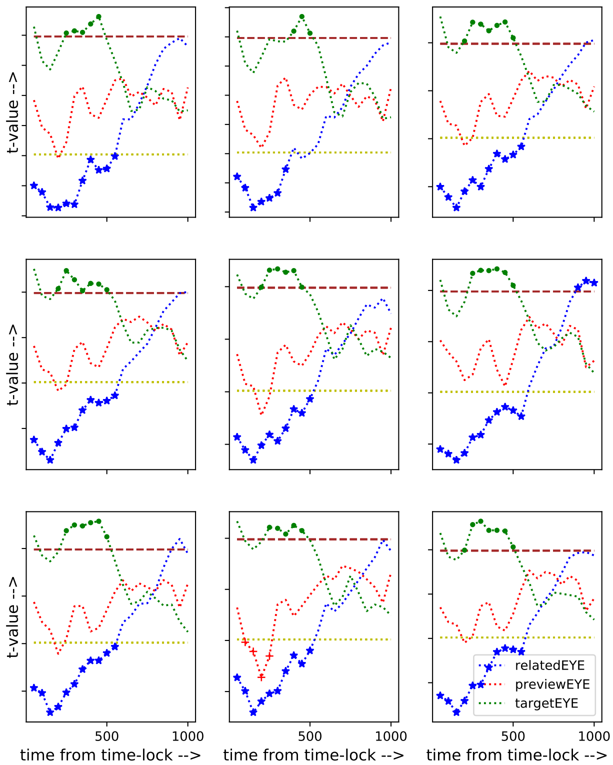

Figure 2: The t-values of the eye track features for the semantically related words.

While the eye track features do not affect the significance pattern of the main factors, it is still interesting to investigate the time course of their own significance, especially in the processing of the related word. The results are shown in Fig. 2, for the model that contained all three features. The first observation is that the patterns are very similar in all nine areas, and that significant t-values are restricted to a time interval that ends at about 550 ms after the onset of the words. Perhaps not surprisingly, in the presence of the eye track features aligned with the two nouns, the preview feature is almost never significant. The eye track features aligned with target and related word are antagonistic: a large value of the eye track feature during the target word yields less negative going ERPs in the related word, while a high value of the feature in the semantically related word is associated with more negative-going ERPs. For the target words (not shown) high eye activity is associated with more negative-going ERPs, again up to 550 ms after word onset. Here, however, there are significant contributions of the eye activity in the preview interval, in the form of less negative-going ERPs in the time interval between 600 and 800 ms after word onset.

4. DISCUSSION AND CONCLUSION

In this study we ask two questions: does a reduced word form in the first part of a sentence affect the recognition of a semantically related word at the end of that sentence differently between native listeners and advanced learners of the language? And does information in eye movements recorded in a visual world paradigm reveal how the visual information is used?

The significant interaction between `reduction`, `language` and `frequency` displayed in Fig. 1 confirms that reduction in the first part of a sentence does affect processing differently in native listeners and advanced learners, and that the difference is larger for high-frequency words. The effect is limited to the first 350 ms after the start of the words, which suggests that it is mainly related to phonetic decoding. We find the same effect of three-way interaction in the time interval up to 350 ms after the start of semantically related words at the end of the sentence. Thus, differences between natives and learners in phonetic decoding of reduced words in the first part of a sentence appear to affect their phonetic decoding of fully-produced related words at sentence end. ERP components in a broad time interval starting around 500 ms after the onset of the target en the related words are usually associated with semantic integration. In that time interval the effects of the three-way interaction differ between the targets and the related words. In the targets the interaction enhances a P600 effect, while that effect is diminished in the semantically related words.

The eye movements of the natives differ from those of the learners in all three time intervals. Therefore, it is safe to conclude that the natives use the visual information differently that the learners. The results shown in Fig. 2 show that eye movements make significant contributions to predicting EEG signals in the time interval up to 600 ms after the onset of the semantically related words. This suggests that visual information affects phonetic decoding and lexical access, but that its role in semantic integration is limited.

5. REFERENCES

[1] S. W. L. Wong, P. P. K. Mok, K. K.-H. Chung, V. W. H. Leung, D. V. M. Bishop, and B. W.-Y. Chow, "Perception of native english reduced forms in chinese learners: Its role in listening comprehension and its phonological correlates," *TESOL QUARTERLY*, vol. 51, pp. 7 – 31, 2015.

[2] D. Nouveau, "Limites perceptives de l'e caduc chez des apprenants nÃ©éerlandophones," *Revue Canadienne de Linguistique Appliquée*, vol. 15, pp. 60 – 78, 2012.

[3] K. Mulder, S. Brand, M. Ernestus, and lou Boves, "Processing reduced speech in the l1 and l2: A combined eye-tracking and erp study," *Language, Cognition and Neuroscience*, Submitted, The data used in this study were collected as part of the European Research Council Consolidator grant 284108 awarded to Mirjam Ernestus.

[4] B. New, C. Pallier, M. Brysbaert, and L. Ferrand, "Lexique 2 : A new French lexical database," *Behavior Research Methods, Instruments, & Computers*, vol. 36, no. 3, pp. 516 – 524, 2004.

[5] B. R. Manor and E. Gordon, "Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks," *Journal of Neuroscience Methods*, vol. 128, pp. 85–93, 2003.

[6] S. Makeig, A. Bell, T.-P. Jung, and T. Sejnowski, "Independent component analysis of electroencephalographic data," in *Advances in Neural Information Processing Systems*, D. Touretzky, M. Mozer, and M. Hasselmo, Eds. Cambridge, MA: MIT Press, 1996, pp. 145–151.

[7] M. Crosse, G. Di Liberto, A. Bednar, and E. Lalor, "The multivariate temporal response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli," *Frontiers in Human Neuroscience*, vol. 10, 2016, article 604.

[8] L. ten Bosch, L. Boves, and K. Mulder, "Analyzing reaction time and error sequences in lexical decision experiments," in *Proceedings of Interspeech*, Graz, Austria, 2019, pp. 2280 – 2284.

[9] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[10] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022. [Online]. Available: https://www.R-project.org/

[11] K. Winsler, K. J. Midgley, J. Grainger, and P. J. Holcomb, "An electrophysiological megastudy of spoken word recognition," *Language, Cognition and Neuroscience*, pp. 1 – 22, 2018.

[12] A. Nijveld, L. ten Bosch, and M. Ernestus, "ERP signal analysis with temporal resolution using a time window bank," in *Proceedings of Interspeech*, Graz, 2019, pp. 1208 – 1212.

[13] P. M. Alday, "How much baseline correction do we need in ERP research? Extended GLM model can replace baseline correction while lifting its limits," *Psychophysiology*, vol. 56, no. 12, p. e13451, 2019.