Processing Lexical Bundles

Lensink, S.E.

Author: Lensink, S.E.
Title: Processing Lexical Bundles
Issue Date: 2020-06-04

CHAPTER 3

Listening

Processing spoken multi-word units: an ERP investigation

Saskia E. Lensink, Antoine Tremblay, Lilian Ye, Arie Verhagen, Niels O. Schiller

abstract

We studied the on-line processing of auditorily presented lexical bundles. Participants were presented with a set of high-frequency lexical bundles and matched controls, while EEG data was collected. We found a sustained early negativity with an early onset that was more pronounced for the matched control items. The data were analyzed using conditional inference random forest modeling (CForest) to gain detailed insights into the time course of auditory processing of lexical bundles, the possible neural sources recruited over time and linguistic and non-linguistic factors that mediate auditory processing. We propose there are three stages that are reminiscent of single word comprehension, representing 1) predictive and bottom-up processes; 2) inhibition and competition; 3) lexical integration. The data provide evidence for an interactive processing model.

Keywords ERPs, multi-word units, auditory processing, comprehension, conditional forest modeling

3.1 Introduction

There is a growing body of work suggesting that language users are sensitive to phrasal frequencies (Bannard and Matthews, 2008; Shaoul and Westbury, 2011; Siyanova-Chanturia, 2013). This is not surprising when considering idioms such as *kick the bucket*, where the meaning of the single words combined does not equal the meaning of the whole. However, phrasal frequency effects are also found for frequent, completely regular, and transparent combinations of words (Arnon and Snider, 2010; Tremblay and Tucker, 2011). These combinations are often referred to as 'lexical bundles'. Examples of lexical bundles are *on the day* or *I think that*. These combinations are thought to play a role in processing because of their common co-occurrence, which encourages the brain to chunk these words together into building blocks of language (Bybee, 2006; Green, 2017).

There is a lot of experimental evidence that phrasal frequencies play a role in processing when speaking and reading. However, it is not yet clear if these phrasal frequencies are also important when listening to language. Moreover, little is known about the time-course of processing of common combinations of words. This study aims to fill this gap by presenting an EEG study on the on-line processing of auditorily presented lexical bundles.

3.1.1 Previous work on the time course of multi-word unit processing

Previous studies have sought to understand the time-course of lexical bundle processing by running eye-tracking and EEG experiments. The electrophysiological signal of the brain can be used to derive Event Related Potentials or ERPs, which are reflections of on-line processing unfolding over time (Kutas and Van Petten, 1994). Only a small number of studies has used ERPs to investigate frequent combinations of words, and most of those studies focused on idioms. In a reading study, Vespignani et al. (2010) compared ERPs elicited by idioms and literal sentences. Past the recognition point of the idiomatic phrases – the word past which participants could recognize the phrase as being idiomatic – idioms elicited an enlarged P300 as compared to their matched literal phrases. The P300 has been found when participants are presented with highly predictable items, such as the lexical item *white* after the presentation of the sentence *The opposite of black is . . .* , (Roehm et al., 2007), or the correct answer to a simple calculation (Fisher et al., 2010). The presence of a P300 in the Vespignani et al. (2010) study shows that after the recognition of an idiomatic phrase, items completing the idiom are actively predicted and pre-activated.

Two previous studies have looked at ERPs of lexical bundles (Hendrix et al., 2017; Tremblay and Baayen, 2010). Tremblay and Baayen looked at the electrophysiological signal of participants reading regular four-word sequences. The whole-string frequencies of these sequences ranged from anywhere between very low (0.01 per million) to very high (100 per million). The authors found that a higher whole-string probability corresponded to a more negative N1 and a less positive P1. The N1 and P1 are early ERP components occurring just 100 ms after stimulus presentation. As the earliest reported frequency effects of single words occur around 100 ms after stimulus onset (Hauk et al., 2006; Penolazzi et al., 2007; Sereno et al., 1998), Tremblay and Baayen reasoned that it would not be possible for multiple words to be accessed and combined within this short time frame. Therefore, they argued that their results show that four-word sequences are retrieved holistically, as if they were a single word.

A couple of years later, Hendrix et al. (2017) investigated the online processing of lexical bundles by presenting participants with a prime consisting of a preposition plus a definite article, followed by a picture of a concrete object. The prepositional phrases had different phrasal frequencies. The authors found effects of both single word frequencies and phrasal frequencies during the naming of the object. Effects of single word frequencies were already present 95 ms after stimulus presentation and occurred mostly in the left hemisphere. Effects of phrasal frequencies were seen as a sustained negativity over the left hemisphere, with higher frequencies correlating with more negative voltages. Hendrix et al. (2017) argue that the different ERP patterns observed are evidence that words and phrases are processed differently. Note that Tremblay and Baayen (2010) did not find any sustained negativities in their study, but only found more negative voltages in early ERP components for higher frequency lexical bundles. Note, however, that Tremblay and Baayen (2010) looked at reading, whereas Hendrix et al. (2017) focused on speaking.

Another way to study the time-course of on-line processing is by employing eye-tracking methods. Eye-tracking provides an indirect means of investigating in which order parts of words and sentences are processed and which cognitive processes might be involved, the idea being that eyes focus on the item that is being processed, and that the duration of gazes indicates the ease of processing (Just and Carpenter, 1980). Recently, Lensink et al. (submitted) used eye-tracking to study the on-line processing of lexical bundles. In line with previous research, Lensink et al. found that more frequent lexical bundles are easier to process than less frequent ones. Moreover, the authors found evidence that lexical bundle frequencies already play a role in gaze durations of the first fixation, showing the early onset of phrasal frequency effects in reading.

These previous studies had participants reading silently or aloud from a screen. There are several disadvantages to studying reading behavior in this way. People can have different reading strategies, and the researcher has little control over the order of processing of the item presented. There are also pitfalls to studying the production of lexical bundles, where participants either have to first read from a screen, or recall items from a list, before they start to articulate. Some participants start talking as soon as they have identified or recalled the first word, whereas others might wait until they have read or recalled the whole string. These different strategies are likely to originate from different cognitive processes, which in turn have different effects on the way participants produce speech. However, there is an alternative where the researcher can precisely control and track the order in which participants receive the input: listening.

Many ERP studies investigating the auditory processing of speech study the time course of phonological, semantic, and syntactic processing, and the influence of context. This is done by presenting participants with full sentences that are either correct, semantically anomalous, or syntactically anomalous. Semantic errors are mostly reflected in a larger N400, and syntactic errors are reflected in larger left anterior negativities (LAN) and a more positive P600 (Friederici, 2002). Oftentimes an enlarged early negativity is also found. Some researchers interpret this negativity as a marker of a phonological mismatch between what is expected and what is heard (phonological mismatch negativity (PMN); Connolly and Phillips, 1994), whereas others consider it a marker of initial form-based assessment of the incoming signal (N200/250; Hagoort and Brown, 2000; Van Den Brink et al., 2001; Van Den Brink and Hagoort, 2004), or a marker of word category violations (ELAN Friederici, 2002; Steinhauer and Drury, 2012). Besides this early negativity, sometimes a sustained negativity with an early onset is seen in auditory processing (Holcomb and Neville, 1991). Although some of these sustained negativities might originate from spill-over effects due to the processing of the previous word (Mueller et al., 2005; Steinhauer and Drury, 2012), they have been linked to working memory processes in several studies (see e.g. Steinhauer et al., 2010).

3.1.2 Current study

If lexical bundles are used in processing, and if listeners make use of them during listening, then it is expected that there is a difference between the ERPs elicited by lexical bundles and ERPs elicited by matched control phrases. Importantly, this difference is expected to arise from the moment that listeners notice that they are listening to a lexical bundle, instead of just any combination of frequent words.

As soon as a listener strongly suspects she is listening to the first part of a lexical bundle, she will expect to also hear the last word of that lexical bundle. When indeed the utterance continues as expected, this match between the expected and the observed might elicit a P300. However, if a listener hears a different word than expected, his expectations are violated. This, we predict, could lead to a larger N400 component at the unexpected word. The N400 is sensitive to frequency and predictability information and has a more negative amplitude when the frequency is lower or the item is less predictable.

Another possibility is that we will see a slow anterior negativity for infrequent combinations.The amplitude of the slow anterior negativity is thought to reflect the amount of resources devoted to short-term memory processes (Kluender and Kutas, 1993; Steinhauer et al., 2010). Retrieving and combining multiple items from the mental lexicon is likely to require more working memory than retrieving a single item or a lexical bundle directly from memory. This could be be reflected in less negative sustained anterior negativities.

In what follows, we will present our exploratory analyses on the time course of auditory processing of lexical bundles. We will first discuss our methods and materials in Section 3.2. In Sections 3.3 and 3.4, we will present the results and will discuss their implications in Section 3.5.

3.2 Materials and Methods

3.2.1 Participants

We recruited forty Dutch native speakers (ten males, mean age 21.4 years) for this study. All participants had normal or corrected-to-normal vision and none of them reported any hearing deficits. After the experiment, participants received a small financial reward for their time.

Two participants were excluded due to technical issues during the experiment, and two participants were excluded because their score on the Edinburgh Handedness Inventory questionnaire (Oldfield, 1971) was negative, indicating left-handedness. The remaining 36 participants (ten male) were on average 21.3 years of age.

3.2.2 Stimuli

We created twenty-six trigram pairs where we contrasted a high-frequency multi-word unit (MWU) with a matched control (Control). The high-frequency multi-word units consisted of three words and were randomly sampled from a set of 1,000 high-frequency trigrams found in the Dutch *Ten Ten* web corpus (Jakubíček et al., 2013).

Matched controls were made by taking a high-frequency trigram and changing its last word, thereby creating a low-frequency trigram. Crucially, we made sure that, for each MWU and Control pair, the final words had similar frequencies, but that the phrasal frequencies were different by at least a factor ten. To give an example, we used the trigram *een belangrijke rol* ('an important role') and changed its last word to create the trigram *een belangrijke vorm* ('an important form'). These trigrams differ only in their phrasal frequencies, whereas all single word frequencies are similar. This way, we can disentangle the effects caused by the terminal single words and the effects of the phrasal frequencies of the whole trigrams. Similar sets of stimuli have been used in several studies looking into multi-word units (see for example Arnon and Snider, 2010).

To ascertain that our multi-word units had a different phrasal frequency than their matched controls, we checked for prevalence of each stimulus pair in different corpora, i.e. the Dutch *Ten Ten* web corpus (Jakubíček et al., 2013), the *Europarl* corpus (Koehn, 2005) and the *EUR-lex* corpus (Baisa et al., 2016). We extracted the frequencies of the trigrams and their constituent words from the Netherlands Dutch subset of the OpenSonar corpus (Oostdijk et al., 2013). The stimuli and their frequencies can be found in Appendix A.

Besides the target items, we included another 52 trigrams that served as filler items. This resulted in a total of 104 different Dutch trigrams, which we pseudo-randomly put into two different experimental lists. We made sure that there was no semantic or phonological overlap between trigrams within at least two consecutive trials. Furthermore, we took care in inserting at least twenty trials between any MWU and Control pair, to minimize the likelihood that participants would recognize the similar form of *een belangrijke rol* and *een belangrijke dag*. The experiment was built in the experimental presentation software E-Prime 2.0 (Psychology Software Tools).

We recorded a male voice reading out loud the stimulus and filler items using a portable USB 2.0 Audio Interface Quad-Capture UA-55 (Roland) at a sampling rate of 44,000 Hz in mono. We created a list where our MWUs, Controls and fillers were randomly presented. We did not inform the speaker about the intention of our study prior to the recordings. Afterwards, we edited the recordings in Praat (Boersma and Weenink, 2016), adding a 500 ms silence before the onset of each stimulus and scaling all stimuli to an equal intensity of 70 dB. There were no significant differences between the acoustic durations of the control trigrams and the multi-word unit trigrams ($p = 0.6392$).

3.2.3 Procedure

Before the start of the experiment, participants completed a questionnaire on their (linguistic) backgrounds, and they filled in the Edinburgh Handedness Inventory test to check to what extent they were right-handed. All participants gave written informed consent before starting the experimental procedure.

Participants were seated in a quiet and sound-proof room in front of a computer screen. In the room, two audio boxes were placed at the front-left and front-right corner. Answer buttons were present on both armrests of the chair in which the participants were seated. Behind the chair, BioSemi ActiveTwo EEG recording equipment was placed. Participants were all connected to a 32 channel EEG set-up while the experimenter explained the experimental task detailed below.

The experiment started with an instruction screen, which was followed by a short practice block where participants could familiarize themselves with the task, and where the experimenter could check if all audio equipment was working properly. The experiment consisted of a practice block of four trials and two experimental blocks of each 52 trials, separated by a short break. The whole experiment took about ten to fifteen minutes to complete.

Each trial lasted three seconds. It started with a fixation cross that appeared in the middle of the screen for 250 ms. Then, after a silence of 500 ms, a trigram was presented auditorily through the audio boxes. To ensure that participants kept paying attention to the task, one third of the auditorily presented trigrams was followed by a visually presented follow-up phrase. All texts were presented in Courier New, font size 12, in black, on a white background. Participants had to judge whether the follow-up phrase could be a grammatical continuation of the trigram that they had just heard. For a correct answer, they had to press the button on their left, and for an incorrect answer, they had to press the button on their right. The words ‘correct’ and ‘incorrect’ were also printed on the left-hand side and the right-hand side of the screen to aid the participants.

3.2.4 EEG recordings

The EEG was recorded using 32 Ag/AgCl electrodes (BioSemi ActiveTwo), which were placed on the scalp sites according to the standards of the American Electroencephalographic Society (1991). We monitored eye movements with four flat electrodes, two of which were placed above and below the left eye, and the other two were placed to the sides of both eyes. Another two flat electrodes were placed behind the ears, at the mastoids, to monitor jaw movements. We used the CMS and DRL electrodes as our ground reference and sampled the EEG signal at 512 Hz. Afterwards, the EEG signal was re-referenced off-line to the mean of the two mastoids and band-pass filtered (0.05-30 Hz) in Brain Vision Analyzer (version 2.0). Eye blinks were corrected by means of an ICA procedure.

Auditory comprehension studies are known to be susceptible to spill-over effects due to processing differences of the words prior to the target words. Late ERP components of a word, such as the N400 or P600, may cause artifacts in the ERPs of a subsequent word if the words are in close proximity in time. Also, even lexically identical words may differ prosodically and phonetically when they are followed by different words, leading to co-articulatory differences, which in turn could lead to differences in the EEG signal spilling over into the target word (Steinhauer and Drury, 2012).

Recall that the stimulus items only differ in their last words, but that the second word might contain articulatory traces of this last word. To control for these effects, and to also control for any spill-over effects due to the processing of the second word, we time-locked the ERPs to the onset of the last syllable of the second word and performed a 200 ms baseline correction. We choose for the onset of the last syllable of the second word instead of the onset of the second word, as the number of syllables of the second word differed across stimuli, with two-thirds of the stimuli having a one-syllable closed-class word in second position. We also reasoned that any co-articulatory effects would be most perceptible in the last syllable.

3.2.5 Conditional inference random forest analysis

We analyzed the EEG data using conditional inference random forests (CForests). Random forests are a widely used machine learning algorithm that can be used for both categorization and regression analyses. CForests have been gaining popularity in fields such as genetics, epidemiology, and medicine, and, more recently, have been applied to several (neuro)cognitive and psychological datasets (McWhinney et al., 2016; Strobl et al., 2009) and linguistic data (Tagliamonte and Baayen, 2012).

A random forest algorithm repeatedly splits the data into two groups based on a set of predictors. The first split is made by testing which predictor explains the most variance in a random subset of the data, and then by determining at which level or value of this predictor the subset can be split into two. The algorithm continues splitting the data until it cannot find any significant features that would warrant any further splitting. The result of this first set of steps is a hierarchical structure know as a classification tree.

The name random forest is chosen because the algorithm does not create a single classification tree, but a large set of classification trees, each based on a different random subset of the data. The final model is based on an average of the predictions of the forest. See Strobl et al. (2009) and Hothorn et al. (2006) for a more detailed discussion.

There are several advantages to using CForests over more traditional parametric methods such as mixed-effects regression. CForests are non-parametric models that do not assume that the data follows a specific distribution. They can model any type of (non)linear relations between predictor variables and outcome variables and are very robust to noise. Another significant advantage is that most of the modeling process is data-driven instead of dependent on human decisions. The modeler does not need to define in advance what shape the functional relation between the predictors and the outcome variable has, nor does she have to define which interactions have to be tested. Any strong simple or complex higher-order interactions that are present in the data will be picked up by the model itself.

Whereas the results from both forward and backward model fitting in regression modeling are notoriously susceptible to the order in which predictors are added or deleted (Strobl et al., 2009), CForests do not suffer from this drawback as they represent an aggregated average of a diverse set of classification trees - each of which is built on a random subset of the data and can therefore take on any type of form. It is important to keep in mind that CForests are truly random models in that there might be slightly different results every time the model is run. Stability and robustness are established by growing a large set of trees. Small effects that might go undetected in parametric regression methods, could still surface in some of the trees in the forest and appear in the aggregated results.

In this study, a total of 10,000 trees was grown on random subsets of the data, and furthermore *variable preselection* was applied, where not only each tree is grown on a subset of the data, but each tree node is split using a random subset of the predictors. Variable preselection produces an even more diverse set of trees (Breiman, 2001). The random subsets consisted of 33.2% of the data for each tree, and for each tree, one variable was randomly selected at each tree node.

3.3 ERP results

In Figure 3.1, the ERPs measured at frontal, central, parietal, and occipital electrodes are plotted. At the frontal, central, and parietal electrodes a clear P100 component is visible, which is more positive for multi-word units at frontal and central electrodes at the midline and right hemisphere. The plots furthermore reveal a small N200 component which is most pronounced at frontal electrodes. Hagoort and Brown (2000) report on an N250 component elicited by semantically anomalous words in spoken sentences. Although we did not use anomalous words in the control items, the final word of the control items is less expected and seems to elicit the same type of response, perhaps less strongly, as semantically anomalous words.

Overall, there is a slow-going anterior negativity that seems largest at frontal areas, and which is more pronounced for control items. This negativity is similar to the sustained anterior negativities found in previous studies on the processing of continuous speech (Hagoort and Brown, 2000; Holcomb and Neville, 1991), which have been linked to increased working memory demands as a result of syntactic processing (Coulson et al., 1998; King and Kutas, 1995; Müller et al., 1997). Moreover, research has shown that the distribution of the negativities is wider and less lateralized for increased working memory conditions

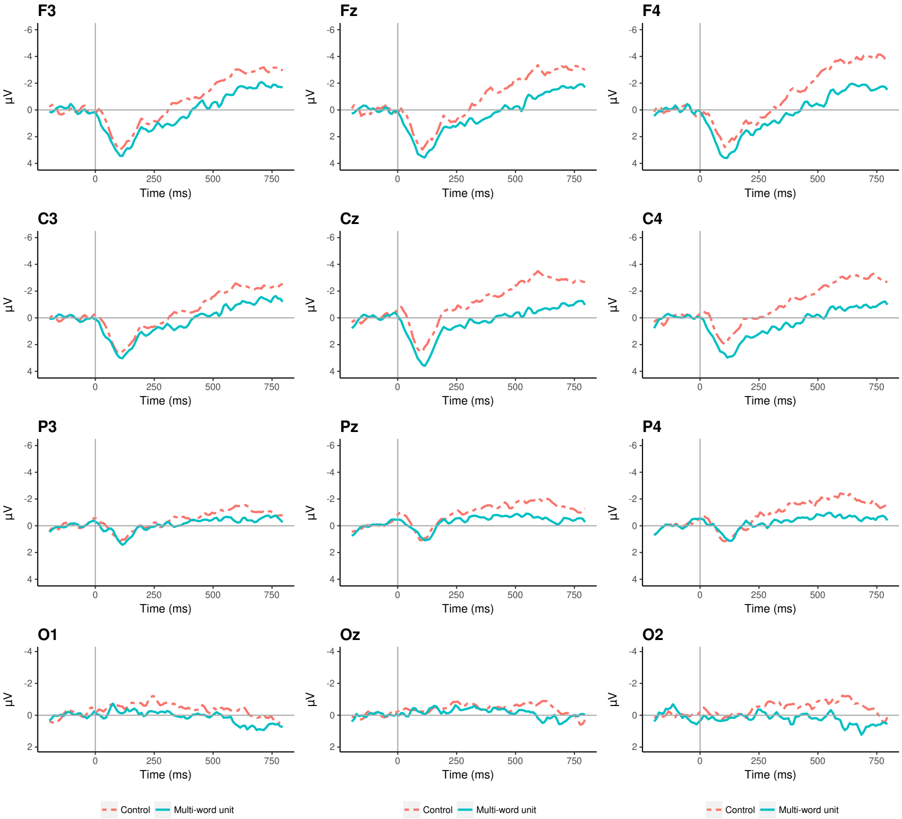

Figure 3.1: Grand-averaged ERP waveforms time-locked to the last syllable of the second word. The control condition is presented in red with a dashed line, and the multi-word unit condition is presented in blue. Presented are frontal, central, parietal and occipital electrodes at the midline, and the left and right hemisphere. By convention, negativity is plotted upwards.

than for grammatical violations (Martín-Loeches et al., 2005), suggesting that our results reflect increased working memory demands when participants are listening to control items as opposed to high-frequency multi-word units.

Multi-word units and control items start to diverge very early at frontal and central electrodes, with a divergence already visible at the P100 at the midline and right-hemisphere. At the parietal and occipital electrodes, the conditions start to diverge from approximately 200 ms after the last syllable of the second word. Overall, the divergence is widely distributed, and seems most prominent at fronto-central regions. Our results are the opposite of the pattern demonstrated by Hendrix et al. (2017), who found more negative voltages for high frequency phrases, which was moreover most prominent in parietal and occipital regions. Note, however, that Hendrix et al. (2017) used a production task. Moreover, the authors used a picture naming task, which needs involvement of the visual cortex, which could explain the more posterior distribution of their results.

3.4 CForest modeling results

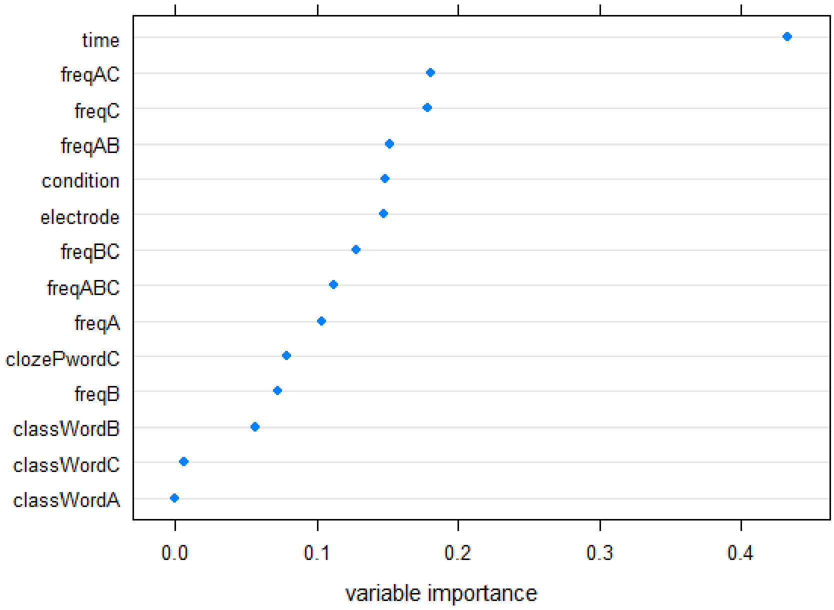

Figure 3.2: Dotplot that shows how much each predictor contributes to explaining the variance in the electrophysiological signal.

Even though interpreting an opaque model such as a conditional random forest is not straightforward, it is possible to study which variables are considered most important by the model, and to study which patterns emerge in individual trees. Figure 3.2 shows how important each variable is within the whole conditional inference forest model in describing the data. The higher on the list, the more variance in the data is explained by that variable.

The variable time is by far the most important predictor, which determines to the largest extent what voltage is generated by the pyramid cells. As the signal fluctuates quite a lot over time, this result is not surprising. A bit lower on the list is the electrode, showing that the location on the scalp also determines to a moderate extent how the signal is manifested. This is not surprising either as EEG data typically shows a lot of variation between different parts of the scalp.

The next items on the list are more interesting: Clearly, the skigram and last bigram frequencies (freqAC and freqAB), the last word frequency (freqC), and the condition (multi-word unit or control item) play the largest roles in explaining the shape of the signal. It is interesting to see that the frequencies of constituent two-word combinations have a larger impact than the frequencies of the three-word combinations themselves.

The the last bigram, the full trigram, and the first and second word frequencies (freqBC, freqABC, freqA, and freqB), the cloze probability of the last word and the class of the second word play a moderate role, whereas the word classes of the first and third words are quite small.

Although these variable importance plots can be insightful, they cannot tell us anything on the direction of an effect — i.e. does a higher bigram frequency correlate to a more negative or a more positive signal? —, nor does it show which interactions might exist — i.e. from which moment in time does the trigram frequency start to play a role? And are any frequency effects located in a specific region?

Because we are interested in learning how lexical bundle processing proceeds over time, it is worthwhile to dive into the structure of a conditional inference tree, to get a better grasp of how the different predictors interact over time. This is a trade-off: When only looking at the complete model, one has to accept that it is quite opaque and cannot provide us with in-depth and detailed insights. It will however be a quite accurate model for making predictions. However, by also looking at only a part of the model, one can conduct a detailed and in-depth study of what factors play a role, how they interact, and how they develop over time. In this article, we opt for the latter option, explicitly so to generate new hypotheses on how spoken lexical bundle processing proceeds, to guide future research.

3.4.1 Results Conditional Inference Tree

In Figures 3.3, 3.4, 3.5 and 3.6, a graphical representation of a conditional inference tree is presented. The tree-like representation shows a large number of binary splits that produce several subgroups of data. Note that all intermediate and terminal nodes have been numbered for easy reference in the text.

The terminal nodes at the bottom of the figure represent the different sets the data has been grouped into by the model, with the number of data points and the average amplitude in microvolts of that specific group. For example, the leftmost terminal node, number seven, consists of 1970 data points (n = 1970), and its average voltage is -0.203 microvolt (y = -0.203). The binary splits are displayed in order of importance: The highest split, node number one, splits the data into two different time bins, and constitutes the strongest predictor of amplitude values. It is interesting to see that the most important binary split happens at 402 ms. The N400 component is widely reported to take place around this time, and has been connected to processes of lexico-semantic integration (Kutas and Van Petten, 1994; Steinhauer et al., 2008).

The subset of data generated before 402 ms can be found at the left-hand side of the figure. This subset, in turn, has been split into a group of fronto-centro-temporal electrodes and more parieto-occipital electrodes by node number 2. Node number 3 splits the subset of fronto-centro-temporal electrodes in a subset of data generated before 270 ms and a subset of data generated between 270 and 402 ms after the onset of the last syllable of the second word. The further one looks down the tree, the more interactions become apparent, and the way in which different factors play a role in different subsets. Therefore, the CForest analysis allows for an in-depth investigation of which factors play a role at different stages of lexical access.

In what follows, we will discuss the results chronologically and topographically, focusing on which factors play a role at different time windows and at different locations. We will relate the modeling results to what is known about the timing of lexical processing and the possible neural sources of subprocesses of lexical access. Although the source of an ERP amplitude is hard to establish on the basis of the mere location of the signal (Steinhauer et al., 2008), it is useful to speculate on its possible neural sources and what the cognitive functions of these possible sources can tell us about how processing proceeds.

As the model clearly subdivides the data into three periods, we propose that these subsets reflect three stages of lexical access. Stage 1 takes place up until 270 ms after hearing the onset of the last syllable of the second word of the trigram. Here, participants predict what might be coming next, while at the same time making use of bottom-up information. At stage 2, taking place between 270 ms and 402 ms, processes of inhibition and competition start to play a role. Then finally, at stage 3, which takes place after 402 ms, lexical integration of both bottom-up and top-down information takes place.

3.4.2 Stage 1: Prediction and bottom-up information, 0-270 ms

Stage 1 starts when participants hear the last syllable of the second word of the trigram. Since this syllable is likely to already contain acoustic cues of the

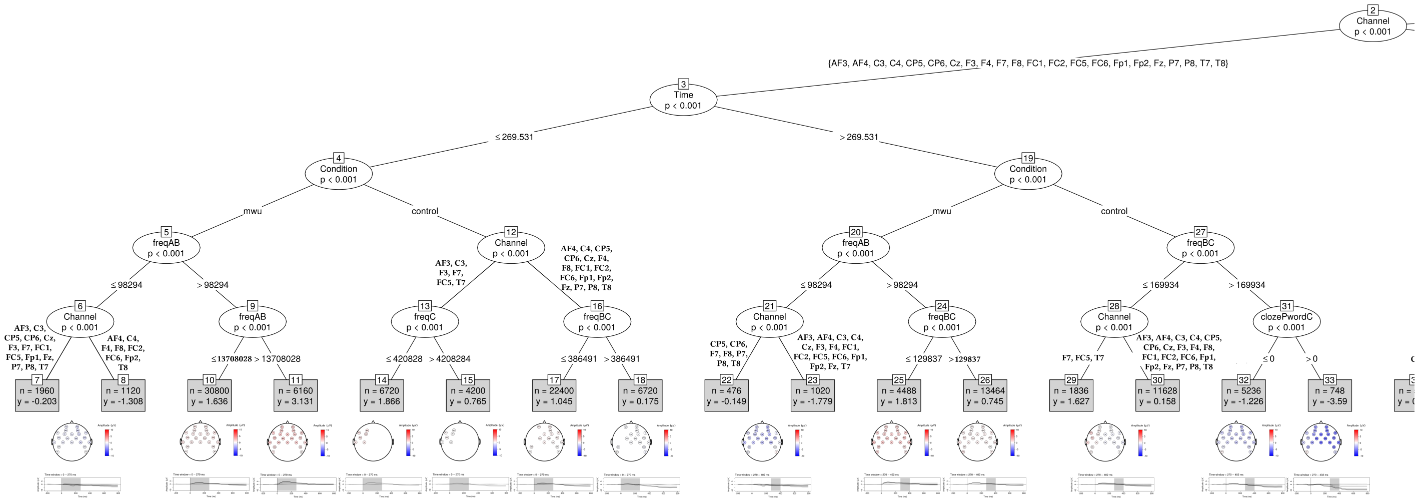

Figure 3.3: Parts of stage 1 and 2 of multi-word processing, visualized here as the first part of a representation of the model-predicted global amplitude values as a function of time, location on the scalp, condition, length of the trigram, unigram, bigram and trigram frequencies, last word cloze probabilities, and word class of the single words included. Global amplitude values are depicted at the terminal nodes, and by means of the colors of the electrodes on the scalps below the terminal nodes. Below the scalps, a time line is provided with the time window concerned highlighted.

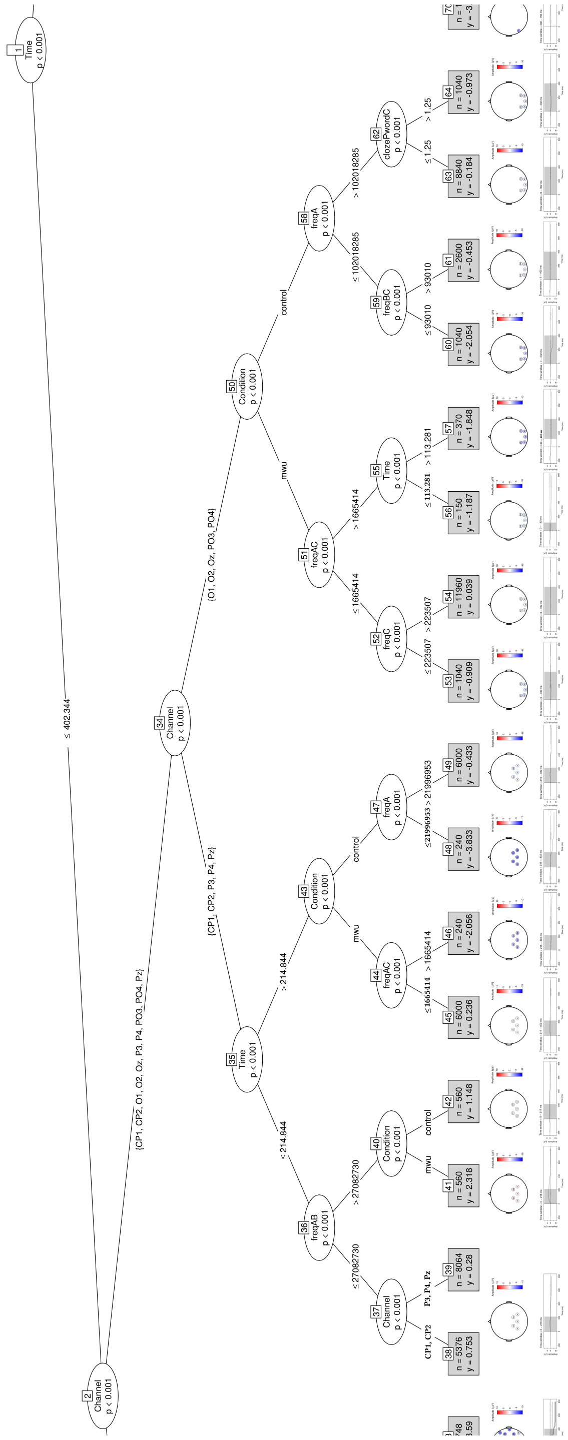

Figure 3.4: Second part of the CForest model, showing what happens at the first and second stages of multi-word unit processing.

following word, we hypothesized that this could be the earliest point in time at which participants notice a difference between multi-word units and matched controls. And indeed, the modeling results already show different neural responses to multi-word units and matched controls between 0 and 240 ms. In short, it appears that participants have already built up expectations on what to expect next, and have different processing strategies for when the bottom-up information matches their expectations (i.e. they encounter the last word of the multi-word unit they were expecting) or when it violates these expectations.

The presence of these different processing strategies suggests that within 270 ms after hearing the first acoustic cues of the last word, participants are sensitive to properties of the full trigram. As can be seen in Figure 3.3, node number three divides a set of frontal, central and temporal electrodes into a time frame before and after 270 ms, and node four subsequently splits this subset by condition, resulting in a subset of multi-word units and a subset of control items. The model has furthermore split the data into different sets of electrodes, which correspond to a fronto-central region, a centro-parietal region, and a parieto-occipital region. In what follows, we will discuss what happens in each of these different regions at stage 1.

Fronto-central processing (nodes 7-18)

At fronto-central regions, processing spoken multi-word units is mostly influenced by the frequency of the first bigram, freqAB, whereas processing of control items is mostly influenced by the frequencies of the last word and the last bigram, freqC and freqBC. We suggest that this reflects different processing strategies when encountering expected or unexpected lexical items, where an expected item prompts the system to further process the first part of the trigram (freqAB), whereas unexpected items prompt the system to shift its attention to these unexpected, new items (freqBC and freqC).

Recall that the ERPs have been time-locked to the onset of the last syllable of the second word, to take into account co-articulatory cues on that syllable. When hearing these cues, participants are able to infer what the last word of the trigram might be. They have already heard and processed most of the first bigram, and have built up expectations as to which word to expect next. Apparently, hearing cues for a word that completes a high-frequency trigram causes participants to continue processing the first part of the trigram. The last part is predictable, and bottom-up processing of the last part is postponed until a later stage. However, upon hearing cues for a word that does not complete a high-frequency trigram, participant's attention is focused towards the last part of the trigram.

When zooming in further into the processing of multi-word units, we see that higher first bigram frequencies are the most important predictor for amplitude values before 270 ms, and that higher AB frequencies correlate with more positive amplitudes. This likely reflects a reduced N1 and an enhanced P2 component. When considering the ERP plots in Figure 3.1, these early components are indeed visible, with early onsets and most prominently in bilateral fronto-central regions. In line with our findings, Sereno et al. (1998) found early frequency effects at the N1 and P2 components in a lexicon decision experiment, with high frequency items corresponding to more positive amplitudes. In later studies, Sereno et al. (2003) and Hauk and Pulvermüller (2004) also found frequency effects roughly 150 ms after stimulus onset, with again more frequent items eliciting more positive amplitudes.

For the control items, the model has split the data into two regions: A region in the left hemisphere (nodes 14 and 15) that processes the last word, and a region that is mostly located in the right hemisphere, with electrodes in frontal, central and temporal regions (nodes 17 and 18), that processes the last bigram.

Nodes 14 and 15 represent early processing of the last word. We propose that upon hearing the last word of a control item, participants are prompted to first process this new and unexpected information, before they can integrate it into the previous context. Nodes 14 and 15 show activations in a region that could originate from the left primary auditory cortex (PAC), where auditory processing takes place, or the left inferior frontal gyrus (LIFG), an area that has been connected to linguistic processing in a wide range of studies (see Vigneau et al., 2006, for a meta-analysis of language processing in the left hemisphere). If we follow Friederici's (2012) proposal for a cortical language circuit for auditory processing, then we expect this bottom-up input to pass from the auditory cortex to the anterior superior temporal cortex and then to the prefrontal cortex.

Nodes 17 and 18 of the model represent early processing of the last bigram. The model shows more positive amplitudes for control items with low second bigram (freqBC) frequencies, in mostly right hemisphere regions. The right hemisphere has been implicated in context processing and general attentional and working memory processes (Vigneau et al., 2011) and is claimed to be biased toward bottom-up, more post hoc, interpretive processing (Federmeier, 2007). Considering the relatively large portion of the right frontal hemisphere that is significantly activated in these subsets, it seems plausible that these activations reflect attentional processes and the recruitment of working memory, where the second and third word of the control item are considered at once. Moreover, bilateral peaks in the temporal lobes have been found to be activated during sentence comprehension tasks, and more specifically by tasks where participants had to generate the last word of a sentence (Kircher et al., 2001; Vigneau et al., 2011). As participants were at this point listening to the last part of a trigram, it is likely that they recruit this region associated with sentence completion tasks.

Centro-parietal processing (nodes 38-42)

Before 214 ms, amplitudes at electrodes CP1, CP2, P3, P4 and Pz are more positive for higher first bigram frequencies. When this frequent first bigram is also part of a multi-word unit, then amplitudes are even more positive. This is similar to the activations seen in the fronto-central regions (see above). Moreover, more positive amplitudes in an early time window for higher frequency items has also been found in previous studies (Hauk and Pulvermüller, 2004; Sereno et al., 1998, 2003).

Parieto-occipital processing (nodes 53-64)

As in fronto-central regions, the most important predictor in this region is condition, which shows that multi-word units are processed differently from controls items in more posterior regions too, and already at an early stage. Note that the terminal nodes of this subgroup are not part of the subgroup of data that has been split into time windows before and after 270 ms — rather, these terminal nodes represent what happens in the centro-parietal regions between 0 and 402 ms after participants heard the first signs of the terminal word of the stimuli. Generally, language processing does not seem to take place in the occipital lobe (Friederici, 2012). However, as EEG is quite imprecise in terms of localization of the neural source, it is possible that the activations seen in the occipital regions originate from more parietal regions.

Multi-word unit processing at parieto-occipital regions is influenced by the frequencies of the last words (freqC) and skipgrams (freqAC), with higher last word frequencies correlating with more positive amplitudes, but with higher skipgram frequencies correlating with more negative amplitudes. More positive amplitudes for higher frequencies also occur in fronto-central and centro-parietal regions. However, it is surprising to see more negative amplitudes elicited by more frequent skipgrams.

The more negative amplitudes for higher skipgram frequencies are unexpected, as the results discussed above all show more positive amplitudes for higher frequencies in multi-word units. Pylkkänen et al. (2004); Tremblay et al. (2016) reported increased activity in their MEG studies around 350 ms (M350) in response to higher lexical and n-gram frequencies. This M350 has been linked to lexical access and indexes inhibitory neural responses. It is possible that high skipgram frequencies cause the listener to consider alternative trigrams, leading to enhanced competition from similar forms, which in turn could lead to increased processing costs as reflected in the more negative amplitudes.

In general, the early posterior activations might reflect the first stage of combinatorial processing and the integration of the last word with the rest of the trigram. In their MEG study investigating language networks involved in n-gram processing, Tremblay et al. (2016) report on a network that is mainly located in posterior regions (their 'Network 3') and whose main function seems to be integrative processing of several sources of information. The network includes areas associated with sentence processing, semantic and discourse coherence processing, and the integration of complex semantic and syntactic information (among others the posterior superior temporal sulcus and the angular gyrus) (Friederici, 2012; Vigneau et al., 2006). As such, it seems probable that the early activations elicited by last word and skipgram frequencies originate at an integrative network located at posterior areas.

As for control items, we see more negative amplitudes for lower first word (word A) and lower second bigram (bigram BC) frequencies. If the second bigram has a high frequency, however, amplitudes tend to be less negative. This pattern is similar to what we saw in fronto-central regions, where lower frequencies also correlate with more negative amplitudes. Note that we also see an influence of the first word frequency for control items, which we have not seen in more frontal and central regions.

When the first words of a control item has a high frequency, there is also an interaction with the cloze probabilites of the last word of the control item: High cloze probabilities of the last word correlate with more negative amplitudes. This is unexpected, given that this subset of the data is within the time range of the N400 component, and higher cloze probabilities have been found to correlate with a smaller, and thus less negative, N400 component (Kutas and Van Petten, 1994). Moreover, Penolazzi et al. (2007) found more positive-going amplitudes between 280-320 ms at posterior mid-line electrodes for high probability words than for low probability ones.

As might be the case with higher skipgram frequencies correlating with more negative amplitudes, we suspect that the more negative amplitudes for higher cloze probability last words index greater processing costs as a result of inhibitory processes. Once a listener has realized that s/he is not listening to a frequent multi-word unit, s/he is not expecting to hear words with high cloze probabilities, as these are more likely to occur in multi-word units but not in control items[1]. S/he will therefore actively inhibit words with high cloze probabilities. This extra inhibition will make it harder to identify a high cloze probability item.

Discussion Stage 1

In general, in fronto-central regions, low-frequency items elicit more negative amplitudes and high-frequency items elicit more positive amplitudes, with multi-word units eliciting more positive amplitudes overall. The more positive amplitudes of the multi-word units seem to reflect reduced N1 and enhanced P2 components (Sereno et al., 1998, 2003; Hauk and Pulvermüller, 2004). Processing spoken multi-word units is mostly affected by first bigram frequencies, whereas listening to spoken control items is mostly affected by second bigram and last word frequencies.

These differences in processing suggests that listeners engage in predictive processing when they encounter lexical items that could form the beginning of a lexical bundle. If their top-down expectations match the subsequent input, listeners continue processing the first bigram, engaging in lexical selection and

[1] However, this is not necessarily the case. A trigram can have a low phrasal frequency, but a high cloze probability last word. Still, this is less common and therefore less expected.

perhaps also lexical integration of that first bigram. However, if their expectations do not match the bottom-up input, listeners focus their attention towards the unexpected input and start processing the last word and the last bigram, engaging in the first stage of spoken word recognition, lexical access.

Posterior regions are also involved at this stage, and even seem to be engaged in later processes of spoken word recognition, i.e. combinatorial and integrative processing. These regions are involved in the processing of skipgrams and the last words of multi-word units, and the first words of control items. Moreover, the cloze probability of the last word of control items also plays a role in these regions. Although the activations could have originated from the primary auditory cortex and therefore only reflect the first stage of auditory processes, it seems likewise plausible that the activations also reflect the involvement of the posterior superior temporal sulcus and the angular gyrus, which perform combinatorial and integrative processing (Friederici, 2012; Vigneau et al., 2006).

It is not only because of the possible neural source that we suspect that the activations seen in posterior regions index later stages in spoken word recognition; it is also because of patterns of activations we observe. Higher-frequency items that correlate with more negative amplitudes likely reflect higher processing costs. Different processes seem to underlie these enhanced processing costs: For higher frequency skipgrams, the larger costs are likely a result of lexical competition effects, and reflect a form of neighborhood density effects (Luce and Pisoni, 1998). For higher cloze probability words (words C of control items), the larger costs are likely a result of inhibitory effects. Most control items are likely to end in a low cloze probability word, and as soon as a listener is aware that s/he is listening to a control item, s/he seems to actively inhibit items that s/he is not expecting to hear, i.e. words with a high cloze probability.

3.4.3 Stage 2: Inhibition and competition, 270-402 ms

The most important data split in stage 2 is type of n-gram: Multi-word unit processing is influenced mostly by the first bigram frequencies (freqAB), whereas control item processing is influenced mostly by the second bigram frequencies (freqBC). In stage 1 there are already some forms of inhibitory and competitory processing in posterior regions. These processes continue and become more prominent and widespread in stage 2. Higher frequency second bigrams in both controls items and multi-word units elicit more negative amplitudes, reflecting larger processing costs: In multi-word units high frequency second bigrams seem to prompt processes of competition between similar forms, whereas high frequency second bigrams in control items are unexpected and prompt processes of inhibition.

Fronto-temporal processing (nodes 21-33)

For multi-word units, lower AB frequencies correlate with more negative amplitudes at fronto-temporal locations between 270 and 402 ms, similar to the direction of the effect before 270 ms. Hagoort and Brown (2000) found a large negative shift around 250 ms for semantically anomalous words, with a mostly central distribution. The authors hypothesize that the N250 might reflect the lexical selection process that takes place at the interface of lexical form and context integration. Although having lower frequencies is not the same as having a semantically anomalous form, the phenomena are similar in that they constitute less expected events.

For multi-word units with high AB frequencies, the frequencies of the second bigram also start to play a role. The higher the first bigram frequency is, the faster participants will be in processing that trigram, which results in earlier onsets of the processing of its second bigram. High BC frequencies, however, correlate with less positive amplitudes than low second bigram frequencies (nodes 25 and 26). Higher first bigram frequencies elicit more positive amplitudes, whereas higher second bigram frequencies elicit less positive amplitudes. This is similar to the more negative-going amplitudes for high frequency skipgrams, as we saw in the subsection on parieto-occipital processing in Section 3.4.2. Therefore, we suspect that the less positive amplitudes elicited by higher second bigram frequencies in this time window index competitory processes (Pylkkänen et al., 2004; Tremblay et al., 2016).

When considering the control items, there is also a negative correlation between frequencies and amplitudes: Both higher BC frequencies and higher cloze probabilities of the last word correlate with more negative amplitudes (node 33). Once a listener has arrived at stage 2 of processing, s/he has already realized s/he is not listening to an expected multi-word unit, and therefore expects a low-frequency second bigram and a low cloze probability item as the third word. S/he might be actively inhibiting high-frequency bigrams and high cloze probability third words, which could lead to enhanced processing costs for these parts of control items, reflected as more negative amplitudes.

Centro-parietal processing (nodes 45-49)

Between 214 and 402 ms at centro-parietal regions, amplitudes elicited by multi-word units are mostly influenced by skipgram frequencies (freqAC; nodes 45 and 46). Higher skipgram frequencies correlate with more negative amplitudes for multi-word units, possibly reflecting larger processing costs due to enhanced competition from similar forms (Pylkkänen et al., 2004; Tremblay et al., 2016). This is similar to what happens between 0 and 402 ms in parieto-occipital regions (see Section 3.4.2). It seems then, that these competitory processes originate in posterior regions and move forward to (or happen concurrently in) more centro-parietal regions. Control items, on the other hand, are mostly influenced by the frequencies of the first word (nodes 48 and 49), with higher first word frequencies correlating with more positive amplitudes.

Parieto-occipital processing (nodes 53-64)

See 'Occipital-parietal processing' in Section 3.4.2.

Discussion Stage 2

In stage 2, bigrams and words of multi-word units are further processed and integrated. Multi-word unit processing is influenced by the first bigram frequencies and, to a lesser extent, by the second bigram frequencies in frontal and central regions, and by skipgram frequencies (freqAC) in more centro-parietal regions. Processing of control items is influenced by the second bigram frequencies and the cloze probability of the last word in fronto-central regions, and by the frequencies of the first word in centro-parietal regions.

Like in stage 1, multi-word units elicit more positive amplitudes overall. These positive amplitudes seem to reflect a reduced N250 (Hagoort and Brown, 2000). The occurrence of an N250 suggests that at this point in time, lexical selection of the target trigram takes place if the trigram is a frequent multi-word unit. However, if the trigram is a low-frequency control item, then the language system spends more resources on processing the last part of the trigram.

By now, the first signs of the influence of the last words of high-frequency multi-word units start to appear. In fronto-central regions, higher second bigram frequencies elicit more negative amplitudes. These more negative amplitudes seem to reflect enhanced processing costs, which are likely due to competition effects between similar high-frequency bigrams that could complete the first frequent bigram. Similarly, in centro-posterior and in occipital regions, higher skipgram frequencies of multi-word units elicit more negative amplitudes, indexing competition effects of similar skipgrams.

These reflections of competition effects for both bigrams and skipgrams might originate from the angular gyrus and the posterior superior temporal sulcus, locations which have been connected to syntactic and semantic integration and sentence processing tasks (Tremblay et al., 2016; Vigneau et al., 2006). They have moreover been linked to semantic processes at the sentential level (Lau et al., 2008). Therefore, we propose that the emerging activations of the last bigram and skipgram frequencies in these regions reflect integrative processes that link the beginning of the multi-word unit to its ending.

As for control items, we see a continued influence of the last word's cloze probability. Higher cloze probability items are more unexpected in low-frequency trigrams, and as such, elicit more negative amplitudes. Moreover, higher last bigram frequencies also elicit more negative amplitudes, again because the presence of a high-frequency last bigram is unexpected in a low-frequency trigram, which increases processing costs.

3.4.4 Stage 3: Lexical integration, 402-800 ms

At the third stage, the most important split in the data is again made by condition: 402 ms after hearing the first signs of the last word of a trigram,

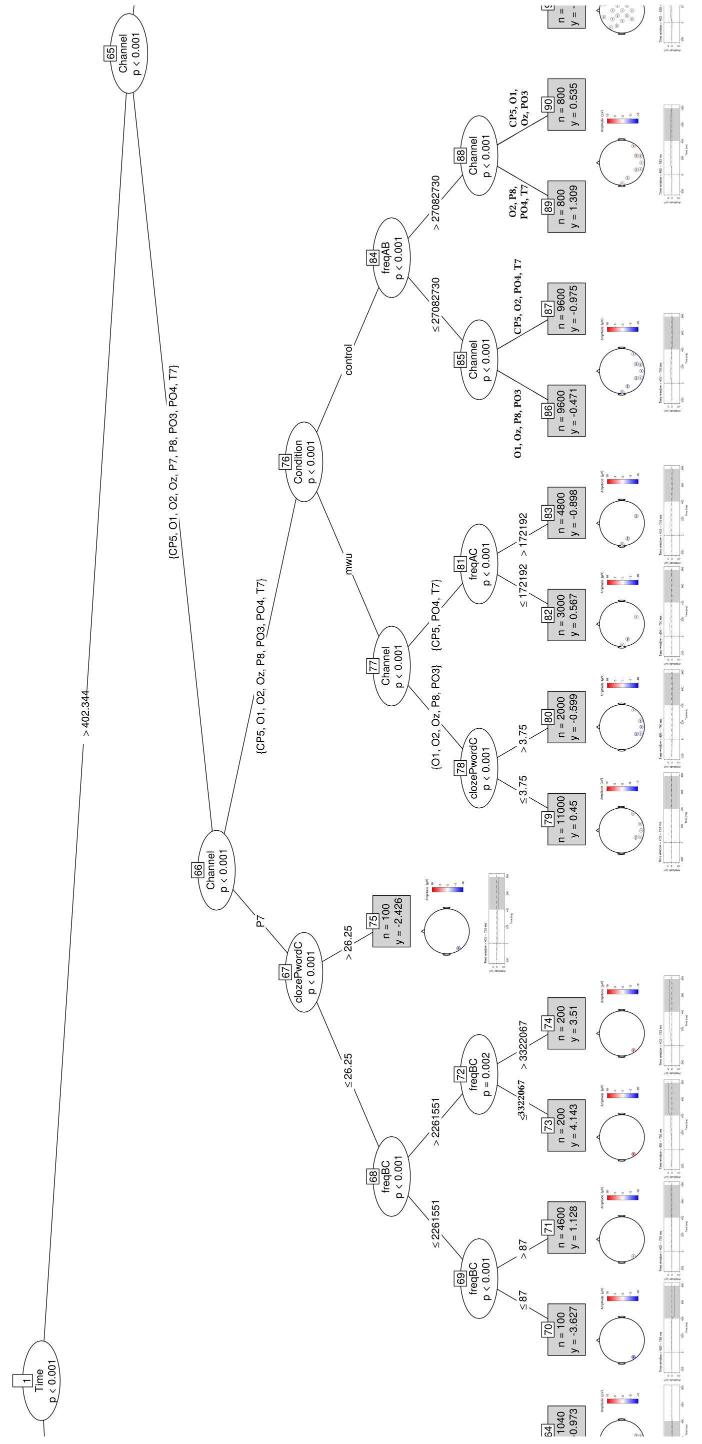

Figure 3.5: Third part of the CForest model, showing what happens at the third stage of multi-word unit processing.

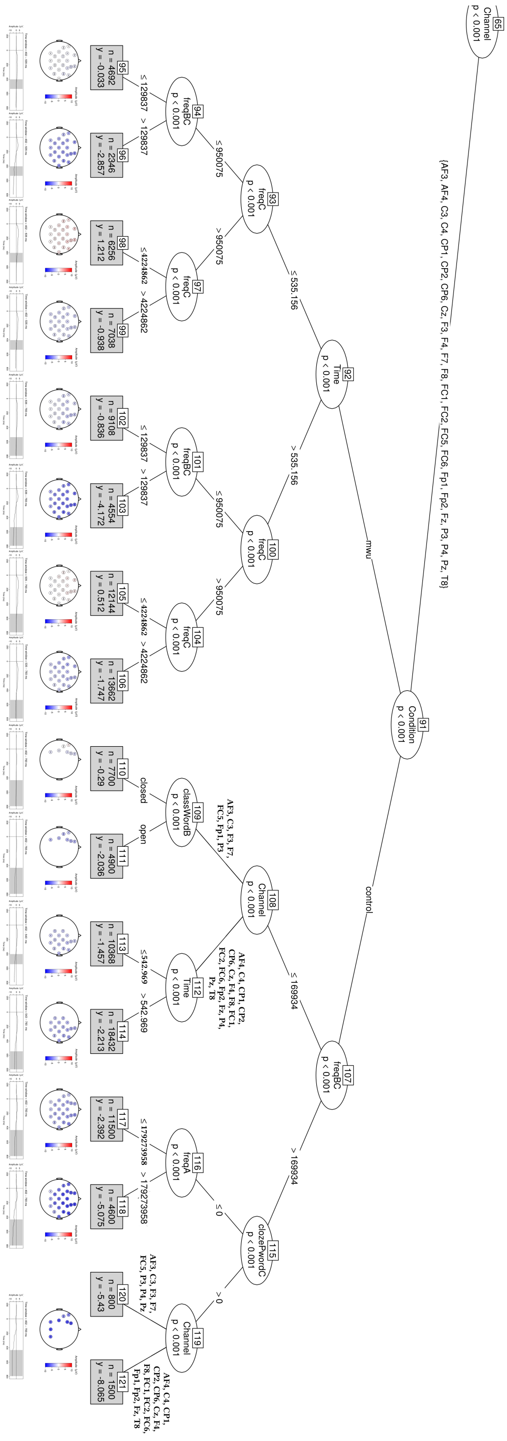

Figure 3.6: Fourth part of the CForest model, showing what happens at the third stage of multi-word unit processing.

frequent multi-word units are still processed differently than matched control items. Specifically, there is a sustained negativity that is less negative for multi-word units and most pronounced in central and right-hemispheric electrodes in fronto-central regions (see also Figure 3.1).

Fronto-central processing (nodes 95-121)

At fronto-central regions, multi-word units are first split by time, into a time window from 402 - 535 ms, and a time window after 535 ms. Both time windows are mostly influenced by the frequencies of the last word, with higher last word frequencies correlating with more negative amplitudes. When the last word frequencies are low, but the last bigram frequencies (freqBC) are high, then amplitudes are even more negative (node 96).

Given that previous research has shown that, for lexical access of single words, lexical integration is taking place after 400 ms (and possibly sooner; Steinhauer et al., 2008), and since we see more negative amplitudes after 400 ms for higher frequency items, we propose that negative amplitudes are indications of easier lexical and contextual integration in stage 3 (Friederici, 2012). This is in contrast to what happens at stage 1 and 2, where more negative amplitudes seem to reflect processes of competition or inhibition.

More negative amplitudes that index ease of processing at this stage 3 are likely to be reduced P600 components. The P600 is an ERP component which is typically elicited by grammatically erroneous sentences, with the incorrect sentence eliciting a greater positivity compared to the correct one. It has been reported to surface as early as 400 ms after stimulus onset (Kaan and Swaab, 2003). Besides grammaticality, the P600 has been reported to vary according to the effort needed to build a coherent syntactic structure (Hagoort, 2003), to reflect continued combinatorial analyses efforts of the brain (Kuperberg, 2007), and to vary according to the degree of probability and salience of a sentence, with more probable sentences eliciting a reduced P600 (Coulson et al., 1998). Frequent BC bigrams and last words of multi-word units are expected, probable, and should therefore take up less processing effort, which would then translate in a less positive, i.e. a more negative amplitude.

In contrast, control items are less likely to be processed as chunks, which means that more combinatorial processes must be at work for control items (Kuperberg, 2007), increasing the P600, leading to more positive amplitudes for control items overall. The most important predictor of amplitude values in these control items is the frequency of the second bigram. Like for the multi-word units, higher frequencies and higher probabilities seem to reduce a P600(-like) component (Coulson et al., 1998). Generally, the higher the second bigram frequency, the more negative the amplitude. If moreover the cloze probability of the last word is also high, then amplitudes are even more negative (node 120 and 121). This reduction in a positive component can also be seen for control items with a low cloze probability last word, as long as their first word has a high frequency; when the first word has a low frequency, amplitudes are more positive (nodes 117 and 118).

When the last bigram of control items is not frequent, amplitudes vary mostly in terms of region on the scalp (node 108). In the left hemisphere, the word class of the second word matters most, with open class words eliciting a more negative amplitude. This more negative amplitude probably reflects ease of processing, as discussed above. As over 90% of the first words of the trigrams start with a closed-class word, most participants will have expected the second word to be an open-class word[2]. When this expectation is violated, a more positive amplitude is elicited (node 110), resembling a P600 effect elicited by unexpected events as discussed by Coulson et al. (1998). In the right hemisphere and central locations, amplitudes vary over time, where amplitudes after 543 ms are more negative. It is likely that these amplitudes reflect context processing, which is known to happen in the right hemisphere (Vigneau et al., 2011).

Parieto-occipital processing (nodes 70-90)

Node 66 separates electrode P7 from the other parieto-occipital electrodes. A low cloze probability of the last word leads to more positive amplitudes at this electrode than high cloze probabilities of the last word (node 75), again showing that, at this third stage, unexpected or improbable events elicit more positive amplitudes. Interestingly, when the cloze probability of the last word is low, and when the frequency of the last bigram is also low, amplitudes are much more negative (node 70). It is not clear why only electrode P7 is split from the other subset of electrodes, and it might be the case that the model is overfitting the data. Note, moreover, that this bin only contains 100 data points, making it unlikely that this effect is robust and generalizable. Future studies could ascertain whether or not this effect is robust.

For the other parieto-occipital electrodes, amplitudes vary by condition. The cloze probability of the last word plays a role in multi-word unit processing in a region of electrodes O1, Oz, O2, P8 and PO3, whereas the skipgram frequencies play a role in multi-word unit processing in a region of electrodes CP5, PO4, and T7. Especially this last region is surprising, as it constitutes a non-continuous region in both the left and right hemisphere. As participants had to judge, at random intervals, whether or not a visually presented fragment could be a correct continuation of the stimuli presented to them, it is possible that this is a prediction network where the skipgram is aiding the lexico-semantic system (CP5, T7) in suggesting possible continuations, which in turn feeds into the visual cortex (PO4) to prepare for a possible visual stimulus. Increased activations in the visual cortex indexing the pre-activation of predicted visual features were also found by Dikker and Pylkkänen (2013).

For control items (nodes 86-90), it is mostly the frequency of the first bigram that plays a role, with higher first bigram frequencies correlating with more

[2] In the top 1,000 trigrams from the TenTenCorpus (Jakubíček et al., 2013), out of which our stimuli have been sampled, 81.6% of these frequent trigrams start with a function word. It seems then, that frequent trigrams tend to start with function words in English.

positive amplitudes. High frequency first bigrams are unexpected in control items, which might explain why more unexpected items elicit more positive amplitudes here, indexing a larger P600 response (Coulson et al., 1998).

Discussion Stage 3

Given the regions involved and the presence of a P600 component, it is probable that during stage 3 lexical integration of all elements of both multi-word units and control items takes place. Moreover, the frequencies of the BC bigrams are playing a clear role in the processing of multi-word units, showing that at this point the last part of the trigrams is also processed and integrated. Whereas higher frequencies correspond to more positive amplitudes for multi-word units in the first two stages, higher frequencies correspond to more negative amplitudes in the third stage. This, we proposed, is likely to be a reflection of a reduced P600 component indexing ease of lexical integration.

The previous two stages involved more positive amplitudes for items that are easier to process. However, at this stage, more negative amplitudes are indicators of ease of processing. A likely ERP component for this stage that reflects ease of processing is a reduced P600. As multi-word units have a higher phrasal frequency, are more expected, and do not necessarily need combinatorial processes, a reduced P600 response is not unexpected. Moreover, higher frequency single words, bigrams, or higher cloze probabilities of items in both multi-word units and control stimuli are also more probable, easier to process, and therefore more likely to elicit reduced P600s — which manifests itself in more negative amplitudes.

3.5 General discussion

We have collected ERP data of participants listening to both multi-word units and their matched controls. We selected a group of high-frequency trigrams and created a set of matched controls by changing the last word for another word that was just as frequent as the original word, but that would not form a frequent combination with the first two words. This way, we could compare processing of high-frequency trigrams and low-frequency similar trigrams that only differed in their last parts.

When a listener encounters a stream of words, at first, s/he cannot know if s/he is listening to a multi-word unit, a low-frequency combination of words, or even a meaningless combination of random words. Because listeners constantly update their expectations based on what they encounter in this stream of words, we expect them to also form expectations on whether they are listening to the first part of a multi-word unit or a low-frequency combination of words [3]. If listeners have different expectations on what to hear next, we expect them to

[3] They will likely not expect a combination of random words, as verbal communication typically carries a meaning and a message.

also employ different processing strategies after hearing at least the first word of a trigram, which should also manifest as different ERP patterns. In other words, we expect to see differences in the ERP data at the moment listeners have already heard and partly processed the first part of a multi-word unit. So to understand if and how spoken multi-word units and matched controls are processed differently, we focussed on the processes taken place after a listener has already listened to the first word and (a part of) the second word of a trigram.

First of all, the ERPs show a clear difference with an early onset between the two conditions. This provides clear evidence for the expectations formulated above, i.e. that spoken multi-word units and matched controls are processed differently. This must be due to the frequency of the combination, as the individual words were matched for their individual frequencies.

Secondly, the different ERPs and their different manifestations provide indications as to what the nature of these differences is: Ease of processing. Overall, we found a sustained negativity that is more positive for multi-word units. Multi-word units show reduced N1 and P2 components, a reduced N250, and a reduced P600 as compared to control items. All these features suggest that multi-word units are easier to process than non-frequent combinations of words (Coulson et al., 1998; Sereno et al., 1998; Hagoort and Brown, 2000; Hagoort, 2003; Sereno et al., 2003; Hauk and Pulvermüller, 2004; Kuperberg, 2007).

Previous experimental work has already shown that ease of processing is manifested as an increase in speed in naming (see e.g. Arnon and Snider, 2010), and greater accuracy in recall (Bannard and Matthews, 2008). In this study, we did not find faster listening per se, but different processing strategies in how multi-word units are processed in comparison to control items. For future studies it will be interesting to explore the possibility that easy of processing in the case of listening to a multi-word unit is also manifested as greater accuracy in processing the auditory signal.

Thirdly and finally, by studying parts of a CForest model, we were able to come up with a detailed proposal on how auditory processing of multi-word units and their matched controls might proceed, and which factors contribute most. For this study, we only focused on the time window where the auditory signal of multi-word units and their matched controls starts to diverge, i.e. from the last syllable of the second word onwards. In view of the early onset of the differences between the conditions in this study, it would be informative to also consider the processing of the first part of spoken trigrams, thereby studying the full course of processing of whole multi-word units.

Our analysis suggests that there are three stages in time during which the last part of either a frequent multi-word unit or a matched control item is processed. The first stage consists mainly of predictive and bottom-up processes, where more positive amplitudes indicate ease of processing. The second stage revolves around combinatorial processes that are influenced by competitive and inhibitory processes, and where again more positive amplitudes indicate ease of processing. The third and final stage consists of integrative processes, where more *negative* amplitudes indicate ease of processing. Units from different levels of complexity play a role in processing, with trigram, bigram, unigram frequencies, and word types of single words playing a role concurrently or in close approximation in time or location. Similar results were found by Tremblay and Baayen (2010), who also found that quadgram probabilities as well as sequence-internal word and trigram frequencies affected event-related potentials.

One of our key proposals is that listeners adapt their processing strategy on the basis of what they expect to hear and what they actually hear — at first focusing more on the first part when hearing a multi-word unit, but more on the last part when listening to a control item — which shows an influence of top-down processes on further processing. These different processing strategies offer evidence in favor of interactive models of auditory processing, where multiple sources of information are employed in parallel (Brink and Hagoort, 2004; Hagoort, 2003; Tremblay et al., 2016).

Acknowledgements We would like to thank Daan van de Velde for his help in creating the stimuli.