Tapping into semantic recovery : an event-related potential study on the processing of gapping and stripping

Ruijgrok, B.J.

CHAPTER 8

ERP experiment IV: Prosodic cues

In this auditory EEG experiment, the character of prosody in relation to the prediction of gapped structure was investigated. I thank Judith Kelholt for her assistance with organising and preprocessing of the stimuli, and recording of the EEG data. I am further grateful to Johanneke Caspers for recording the stimulus material. Cesko Voeten, Jos Pacilly and Olga Kepinska helped me with scripting preprocessing steps for the audio files and EEG analysis.

8.1 Introduction

So far, I have presented data from word-by-word reading tasks. In this experiment, I will investigate Gapping in the auditory modality. As discussed in earlier chapters, it has been suggested that Gapping is licensed by a discourse constraint which is dictated by syntax. Furthermore, prosodic parallelism has been put forward as an important grammatical constraint. I have argued, however, that it may be possible that prosody interacts with interpretation regardless of syntax. In the ERP experiment reported in this chapter, I tested to what extent listeners are guided by prosody during the interpretation of Gapping by comparing conditions that only differ in the way prosody is expressed.

In Chapter 2.5, I asked to what extent a sentence such as (1b), as compared to (1a), is ungrammatical, since the object in the right conjunct lacks an accented counterpart in the left conjunct in (1b). The unaccented object in the first conjunct may be a sign that it will be a candidate for deletion as is the case in (1c).

(1) a. De MAN kocht een BOEK in LONDEN, en de VROUW een KRANT in LEIDEN.
'The man bought a book in London, and the woman a newspaper in Leiden.'
b. De MAN kocht een boek in LONDEN, en de VROUW een KRANT in LEIDEN.
'The man bought a book in London, and the woman a newspaper in Leiden'
c. De MAN kocht een boek in LONDEN, en de VROUW in LEIDEN.
'The man bought a book in London, and the woman a newspaper in Leiden'

It may therefore be hypothesised that the prosody of the first conjunct predicts the remaining structure in the second conjunct. The grammar requires that the object in the second conjunct in (1b) is deleted, if the object in the first conjunct, *boek*, bears no contrastive accent (c.f. the felicity condition on contrastive fragments Griffiths & Lipták, 2014). If, however, this object is accented, as is the case in (1a), the grammar requires a parallel phrase to be apparent in the second conjunct contrasting with its counterpart. In processing terms, this state of affairs could be translated into a cue to predict upcoming (deleted) structure. In other words, a processing difficulty is predicted if an object is deaccented in the first conjunct but accented in the second conjunct. Behaviourally, a processing difficulty may be measured in terms of comprehension scores which may become lower as a function of processing difficulty.

The onset of the determiner of the object determiner phrase in the right conjunct will be taken as measure point to compare ERPs. This is comparable to the replicated ERP experiment reported in Chapter 5.1. If people predict deleted structure, but encounter a remnant instead, I expect a LAN-like component between 200-400 ms may be found on the determiner. In line with Kaan et al. (2013) and Lau et al. (2006), I will regard a LAN as an index of prediction strength (though "index of surprisal" may be more straightforward). In addition, an unexpected noun may elicit an N400 effect, although scepticism is in order given the null effect in the replication study. Additionally, auditory processing is relatively fast, with effects of unexpected items reported to start around 50 ms after stimulus onset (see for example Clementz, Barber, & Dzau, 2002).

8.2 Methods

Test materials

From the replicated study in Chapter 5.1, forty-four stimulus pairs were chosen as exemplified in (2). As explained in Chapter 5.2.2, only sentences with verbal modifiers could be included given that objects should be able to be deleted. The listener may anticipate Gapping of the verb as well as the object in (2b). In other words, this deviant condition can be compared to the control condition (2a) in which an accented contrasting element is available.

(2) a. ANOUK zond de KAART aan haar VADER, en JULIA
Anouk sent the card to her father and Julia
de BLOEMEN aan haar MOEDER.
the flowers to her mother
'Anouk sent the card to her father,
and Julia the flowers to her mother.' (***Parallel prosody***)

b. ANOUK zond de kaart aan haar VADER, en JULIA
Anouk sent the card to her father and Julia
de BLOEMEN aan haar MOEDER.
the flowers to her mother
'Anouk sent the card to her father,
and Julia the flowers to her mother.' (***Non-parallel prosody***)

Since the contrastingly accented object in the second conjunct is unexpected in (2b) as compared to the same object in (2a), an ERP effect may be expected at the determiner preceding the noun. At the determiner a negativity, possibly a LAN, may be expected (as found in the replication study), relating to the violation of a grammatical constraint. The determiner will be unexpected and therefore a LAN may be considered as an index of prediction strength. At the following noun, an N400 may be predicted since the noun is less expected in the context.

In addition to the test stimuli, ten filler sentences containing a coordination with the connective *en* "and" and twenty-two containing a coordination with the connective *maar* "but" were selected from the original data set. The experimental pairs were divided over two lists. An additional five sentences were included as practice items.

All sentences were digitally recorded (44.1 kHz) by a professionally trained Dutch native phonetician in a sound-proof room using a directional Sennheiser MKH-416 condensor microphone. Items were edited and analysed in Praat (Boersma & Weenink, 2017). The duration of experimental stimuli was 4,624 ms on average ($SE = 36.08$) and durations did not differ between conditions [$t(43) = 0.73$, $p = .471$; $M_a = 4{,}650$, $SE_a = 53.40$; $M_b = 4{,}598$, $SE_b = 48.88$]. Representative examples of prosodic contours and their ToDI (Dutch version of ToBI, tones and breaks indices, as developed by Gussenhoven, 2005) transcription are displayed in Figure 8.1.

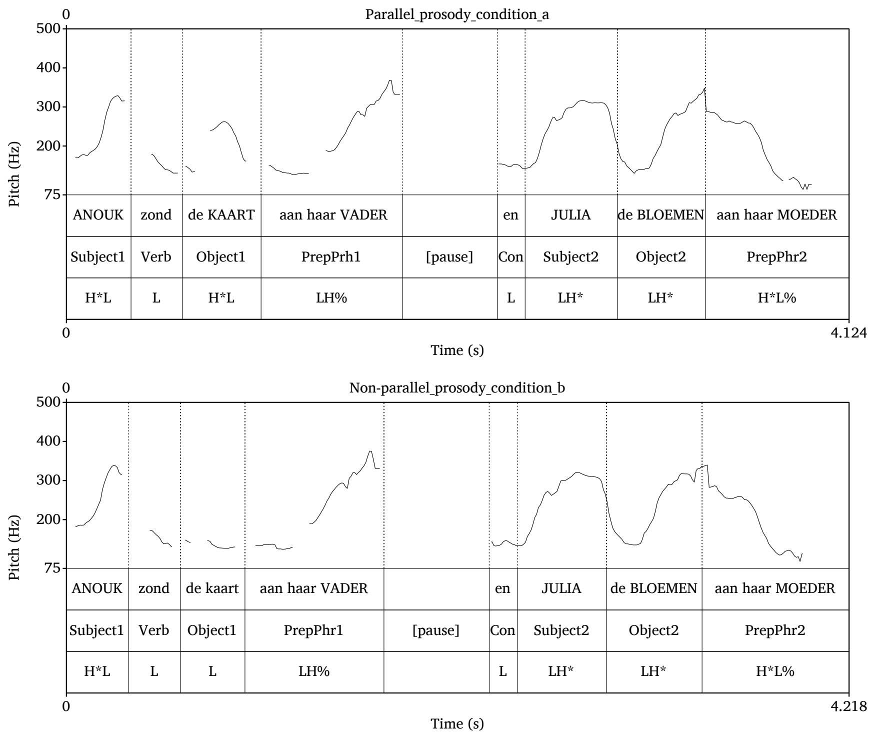

Figure 8.1: Prosodic contours of the Parallel prosody condition (***a***) are shown in the top panel and for the Non-parallel prosody condition (***b***) below. For every phrase, the content is provided, as well as a label and transcription in ToDI.

Each sentence can be decomposed into individual phrases. For every phrase (as depicted in Figure 8.1), the mean of the fundamental frequency and duration were calculated individually. Figure 8.2 represents the relative differences of the fundamental frequency between phrases and Figure 8.3 represents the differences in duration. Additional pairwise t-tests using a Bonferroni correction were performed both to establish differences between and within phrases in terms of pitch and to determine differences in means of duration. The relevant comparisons are listed in Table 8.1.

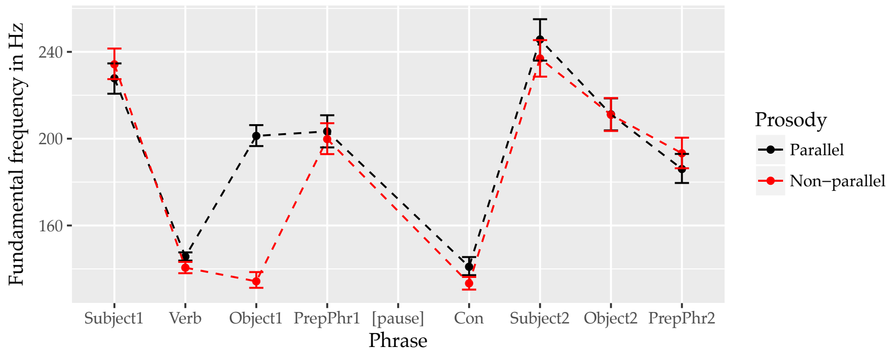

Figure 8.2: Means of fundamental frequency per phrase grouped by conditions of experimental sentences (bars indicate 95% confidence intervals).

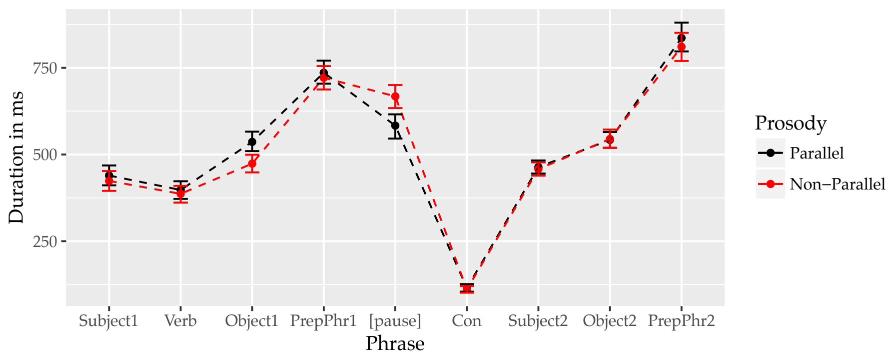

Figure 8.3: Means of duration per phrase grouped by conditions of experimental sentences (bars indicate 95% confidence intervals).

Figures 8.2 and Table 8.1 clearly show the expected differences in means of pitch between the objects of the first and second conjunct. The duration difference between the prepositional phrase of the first conjunct and its counterpart in the second is due to final lengthening. In Dutch, final lengthening of an utterance is longer than final lengthening at a phrase boundary within an utterance (see for example Cambier-Langeveld, Nespor, & Heuven, 1997). A few additional results are notable. In the Parallel prosody condition,

Phrase	Subject				Verb	
Comparison	1a-1b	1a-2a	1b-2b	2a-2b	a-b	
F0	1	.016	1	1	1	
Duration	1	1	1	1	1	
Phrase	Object					
Comparison	1a-1b	1a-2a	1b-2b	2a-2b		
F0	<.001	<.001	<.001	1		
Duration	.425	1	.108	1		
Phrase	PrepPhr				[pause]	en
Comparison	1a-1b	1a-2a	1b-2b	2a-2b		
F0	1	.129	1	1	NA	1
Duration	1	<.001	.003	1	.009	1

Table 8.1: *P*-values of Bonferroni corrected pairwise t-tests. "1" and "2" refer to the first conjunct and second conjunct, respectively; "a" and "b" refer to condition ***a*** (Parallel prosody) and ***b*** (Non-parallel prosody) respectively.

the mean of pitch of the subject in the left conjunct was significantly lower than of that of the right conjunct. In the Non-parallel condition, there was no difference in pitch. Despite this, no significant differences between conditions within phrase position were found. Lastly, it appeared that the relatively shorter durations in the first conjunct in the Non-parallel prosody condition were compensated during the pause which was longer in condition ***b***. As a result, the onset of the determiner in the second conjunct relative to the onset of the sentence was almost equal between conditions [M_a = 3272, SE_a = 0.03; [M_b = 3243, SE_b = 0.03].

Participants

Twenty native Dutch participants with normal hearing (three left-handed, three male, M_{Age} = 23.05, range 19-41) took part in this study and were paid €15. The responses of all participants are taken into consideration in the behavioural data analysis. The EEG data of six participants were disregarded after preprocessing because of too many artefacts. Of the remaining participants (three male, mean age = 23.93, range 19-41) two were left-handed. The experiment followed the Ethics Committee regulations of the Humanities Faculty of Leiden University, which approved its implementation. Participants gave informed consent before the study.

Procedure

The stimuli were presented in a fully randomised order using the presentation software E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA). Stimuli were counterbalanced in such a way that only one item per sentence pair was presented to each participant. Participants were comfortably seated in a dimly lit sound-proof room at a distance of approximately 90 cm from a computer monitor. Each stimulus was preceded by a fixation cross ("+") which appeared at the centre of the screen and remained there for 1,000 ms. Then the audio file was played. After a pause of 1,500 ms a yes/no comprehension question appeared on the screen. The questions referred to different parts of the sentences equally and participants were instructed to answer as quickly and accurately as possible. For half the participants the left response button referred to "YES", for the other half the left button referred to "NO". Between pressing the response button and the next trial, a pause intervened lasting 1,000 ms. After seven trials, a break occurred which could be ended by the participant when they were ready to proceed. Before the actual experiment which contained 76 trials in total, five trials were presented as a practice session.

The experiment was concluded with the working memory test as used in earlier experiments.

The experiment took about 1.5 hours per participant in total, including set-up.

Apparatus and electrophysiological recording

A description of the recording set-up can be found in Chapter 5.1.1.

Data analysis

Using Brain Vision Analyzer Version 2.0 (Brain Products, Munich, Germany), the EEG data were preprocessed before analysis to reduce noise and artefacts as much as possible. Eye blinks were corrected using an Independent Components Analysis procedure (Makeig et al., 1996). Remaining artefacts were rejected on the basis of the same criteria in all earlier experiments reported in this thesis. Epochs of 1,000 ms were computed with a 200 ms pre-stimulus baseline. ERP grand averages were time-locked to the onset of the target word *de*, the determiner of the object in the second conjunct. In total 19.97% of the trials were rejected. Out of 22 trials per condition per participant, 16.93 trials on average in condition ***a*** and 18.29 in condition ***b*** were retained for analysis.

To determine time-windows of interest, a permutation test was carried out at every sample for each electrode with independent factor CONDITION. Considering the small sample, no further statistical tests were run and only an exploratory analysis is carried out.

8.3 Behavioural results

Due to a scripting error one sentence pair was assigned an unrelated question. Therefore, this pair has been excluded from the behavioural analysis. No participants were rejected on the basis of accuracy (M = 96.51%, SE = 0.65%).

The accuracy scores were similar for both conditions ($M_{\text{Parallel prosody}}$ = 96.27%, $M_{\text{Non-parallel prosody}}$ = 96.76%. The difference in mean values was not significant as shown by a paired t-test [$t(19) = -0.39$, $p = .699$, $d = .088$].

The data from the working memory test of one participant were compromised due to a technical malfunction. Therefore, the following analysis is based on 19 participants. The average accuracy score of the three test sessions was 62.46% (SE = 2.87%). Per condition the scores were: $M_{\text{Random Counting}}$ = 56.84%, $M_{\text{Auditory Presentation}}$ = 60.00%, $M_{\text{Visual Presentation}}$ = 70.53%. Although numerically the difference between the random counting and visual conditions seemed large, a repeated measures ANOVA by subjects with CONDITION as independent factor and ACCURACY OF NUMBER RECALL as dependent variable yielded only marginal significance [$F(2, 34) = 2.76$, $p = .077$, $\eta^2_G = .066$].

The scores of the comprehension task of the ERP experiment were compared with the scores of the working memory task (n = 19). A small, non-significant correlation was found between the variables ACCURACY OF SENTENCE COMPREHENSION and ACCURACY OF NUMBER RECALL [$r = .300$, $p = .212$].

8.4 Electrophysiological results

In Figure 8.4, grand averages of Parallel prosody condition ***a*** compared to Non-parallel prosody condition ***b*** at onset of the determiner of the second object are visualised. Average onset of the noun is 141 ms (SE = 0.004) after onset of the determiner. Since an improper baseline can be observed, especially at frontal and central electrodes, further interpretation is not warranted.

In order to identify any between-condition differences that occur before the onset of the determiner, epochs of 1,375 ms were computed with a 575 ms pre-stimulus time-window (relative to the onset of the determiner) and an 800 ms time-window after target onset. The pre-stimulus time-window of 575 ms was determined on the basis of average durations of the connective *en* [$M = 113.00$, $SE = 3.87$] and the subject [$M = 460.98$, $SE = 7.16$] preceding the target. The approximate onset of the connective was taken as the start of a baseline of 100 ms. During determination of longer epochs, 24.84% of the trials were rejected. Out of 22 trials, 6.14 trials on average in condition ***a*** and 4.79 in condition ***b*** were excluded from calculating the grand averages. Figure 8.5 shows the grand averages of Parallel prosody condition (***a***) compared to Non-parallel prosody condition (***b***) of the longer epochs as well as the results of the permutation test ($p \leq .05$) that was carried out for each electrode (by the factor CONDITION).

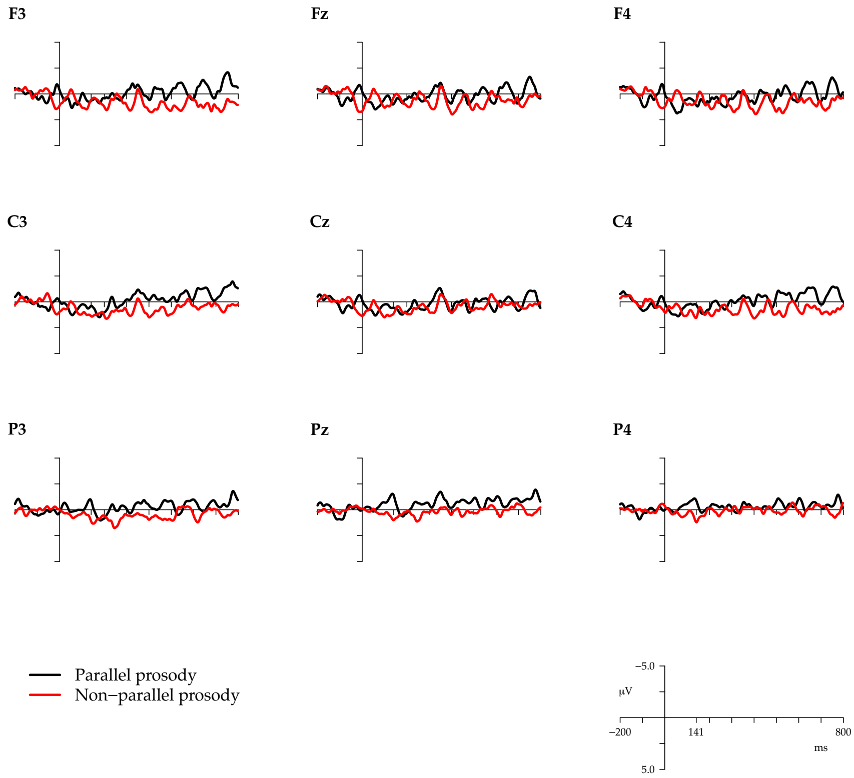

Figure 8.4: Grand averages of Parallel prosody condition (***a***) compared to Non-parallel prosody condition (***b***) at onset (y-axis) of the determiner (*de*) of the object phrase in the second conjunct (*de bloemen*) at electrode sites F3, Fz, F4, C3, Cz, C4, P3, Pz and P4. Approximate onset of the noun (*bloemen*) is 141 ms ($SE = 0.004$) after onset of the determiner. Corresponding example sentences can be found on page 153.

8.5 Discussion

In this experiment I hypothesised that the prosody of the first conjunct predicts the remaining structure in the second conjunct. A LAN-like component between 200-400 ms was expected to be found on the determiner in the Non-parallel prosody condition, reflecting an index of prediction strength. Additionally, I hypothesised that an unexpected noun would elicit an N400 effect, although scepticism was in order given the null effect in the replication study reported in Chapter 5.1. Sentences containing unexpected structures were expected to yield a processing difficulty, as reflected by ERP components, as well

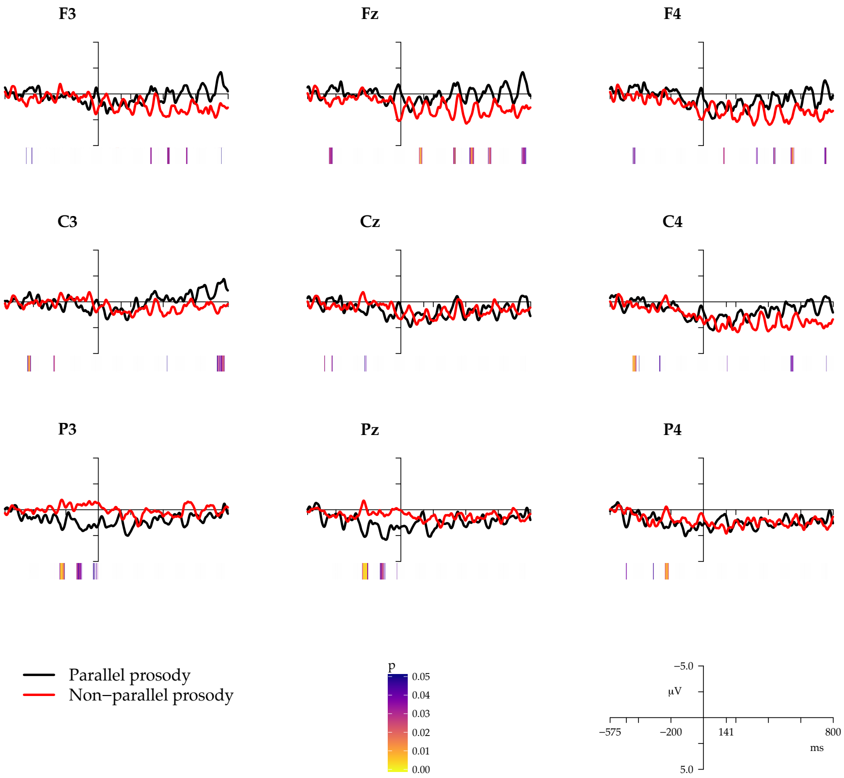

Figure 8.5: Grand averages of Parallel prosody condition (***a***) compared to Non-parallel prosody condition (***b***) at onset (y-axis) of the determiner (*de*) of the object phrase in the second conjunct (*de bloemen*) at electrode sites F3, Fz, F4, C3, Cz, C4, P3, Pz and P4. Baseline of 100 ms starts at -575 ms. Approximate onset of the noun in Object2 (*bloemen*) is 141 ms (SE = 3.74) after onset of the determiner. At each electrode, significant effects ($p \leq .05$) of a permutation test carried out at every sample (by factor the CONDITION) are shown. Corresponding example sentences can be found on page 153.

as by a comprehension scores.

Since the comprehension scores show a ceiling effect, it is difficult to estimate the effect of prosody in terms of behavioural responses. While the questions may have been too easy to answer, it could be that an assumed parallel prosody constraint is lenient enough not to mislead ultimate comprehension. It could also be that such a constraint is not measurable through behavioural means. Despite the ceiling effect, the correlation between the comprehension scores and the working memory test showed the same pattern as in the previous experiments, meaning that low comprehension scores correlate – although not significantly – with low working memory scores.

Although the ERP results are based on a relatively small sample,[1] some effects may be explored and may be corroborated by future research. In this experiment, I have chosen to use naturally recorded stimuli, which means that they were not re-synthesised or edited (except for the onset and offset).

It was hypothesised that modulation of the prosody of the object in the first conjunct would influence the expectations of the processor when encountering the object phrase of the second conjunct. At frontal electrodes, early positive deflections are apparent in the Non-parallel prosody condition starting 100 ms after onset of the determiner of the second object, as well as 100 ms after onset of the object noun. Some ERP researchers assume that P1 components are "wholly or primarily due to the feedforward sweep of activity through the sensory pathway" (Woodman, 2010:2043). With respect to spatial attention on visual processing, a P1 measured at posterior electrodes has been regarded as reflection of increased attention (also sometimes referred to as "selection"; see for a review Hillyard, Vogel, & Luck, 1998) and may be modulated by a participant's state of arousal. This may in turn influence the entire waveform to become more positive beginning with the P1 wave (Luck, 2014:76). A P1 at frontal electrodes has been coined "Frontal Selection Positivity" (FSP) and has also been found mainly in relation to visual stimuli (see for example Michie et al., 1999). In auditory experiments, selective attention has been connected to an early positivity (P50) (Luck, 2014:81). A P50 (peaking between 50-100 ms) may further be modulated by expectancy giving rise to higher amplitudes for unexpected items (Clementz et al., 2002).

Apparently, the deviant prosody condition may have affected the way the processor treats the unexpected object in the second conjunct – giving it more attention. The positive wave may be either extended due to the unexpected item, or – maybe concurrently – higher cognitive components starting at 300 ms reflect processes of reevaluation of the input. Especially at electrode F8 (not visualised), the positivity was significant in time windows of 50 ms and longer, between 300-350 ms, 350-550 ms, 620-670 ms, and 790-880 ms. It should be noted that, apart from attention/selection effects as reflected by an FSP and and a P50, frontal lobe activity has been connected to a wide range of cognitive demands (such as working memory, episodic memory, problem solving, perception), making it difficult to interpret the effects found here as language-specific. However, it may be the case that processes of accommodation, a term used in linguistic theory to refer to mechanisms at the level of discourse and pragmatics (see also Chapter 2.4.1), are connected to these cognitive demands. At best, we could assume that the positivity around 300 ms relates in part to attentive language comprehension – something I have suggested earlier in Chapters 6 and 7.

[1]In contrast to my previous experiments, I included left-handed participants (after visual inspection of grand averages) to arrive at a larger sample; in general, I have come to the conclusion that linguistic research should not be based on data from right-handed people exclusively.

Rather unexpectedly, already at the subject of the second conjunct (starting at approximately 462 ms before the critical measure point) early positive deflections could be seen, mainly at frontal and central electrodes. However, during acoustic analysis of the stimuli it became clear that, in condition ***a***, the means of the fundamental frequency of the subject in the second conjunct differed significantly from the subject in the first conjunct, whereas in condition ***b*** no such difference was apparent. Possibly, the speaker had relatively more difficulty in producing the Non-parallel condition. At the position of subject noun of the second conjunct, only a slight numerical difference between the Parallel and Non-parallel condition was visible. In addition, the pause in condition ***b*** was slightly longer. It is difficult to determine which of the differences could be the origin of the early positivity. Though it is interesting to note that a P50 has been connected to pitch discrimination (Giuliano, Pfordresher, Stanley, Narayana, & Wicha, 2011).

At central and – more prominently – posterior electrodes a negative deflection was visible starting around 200 ms after onset, which may be related to an N2. In visual experiments, a posterior N2 may reflect some aspect of focusing of attention and categorisation processes of a stimulus. In auditory experiments, the mismatch negativity (MMN) peaks around the same time point, yet, it usually has a fronto-central midline distribution (see also Luck, 2014:85). A tentative conclusion could be that the negativity reflects processing of the difference in pitch and possibly the difference in duration of the preceding pause. Again, I would like to stress that future experiments are needed to corroborate these preliminary findings. For example, stimuli could be synthesised manipulating only the object in the first conjunct, neglecting the apparent natural production differences. A useful method is "cross-splicing". Target items can be constructed by cross-splicing the critical segment from the Parallel condition (i.e. the right conjunct) over the corresponding segment in the Non-parallel condition. In another design, conditions could be included in which the object in the second conjunct *is* deleted. In this case I would predict the reverse pattern, as no effects should be demonstrable if the object in the first conjunct is de-accented (c.f. example (1c) on page 152).

All in all, it seems that prosodic cues – if they are ungrammatical – do not necessarily impact full interpretation. In other words, it seems that a prosodic parallelism constraint violation has relatively minor consequences. These consequences seem to be undetectable in behavioural responses, although their effects may be measurable in terms of brain activity.

8.6 Conclusion

In an auditory ERP experiment, I tested to what extent the prosody of the first conjunct affects expectations of ellipsis in the second conjunct. While a processing difficulty was expected for non-parallel prosody conditions, the manipulation of prosody could not be measured by means of a comprehension test. In addition, hypothesised ERP results could not be established due to an improper baseline. Although the stimuli were recorded by a trained phonetician, it appeared that the critical differences in production of the object in the first conjunct may have impacted the speaker's production of words later in the sentence, leading to an improper baseline. However, and despite the small sample, an exploratory analysis showed ERP effects related to attention/selection processes at the predicted time point. Additional ERP effects were established at time points preceding the object in the second conjunct which may be due to the pitch difference of the subject in the second conjunct as well as the duration of the pause between the conjuncts. Future research is needed to sustain the exploratory findings reported in this chapter that suggest that a prosodic parallelism constraint violation has relatively minor consequences.