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

Ruijgrok, B.J.

CHAPTER 7

ERP experiment III: Semantic complexity

This chapter reports an ERP experiment on the difference between determiner *de* "the" and quantifiers *elke/alle* "every/all" in Stripping constructions. I thank Isabella Jordanoska for assistance during data collection.

7.1 Introduction

In the experiments in the previous chapter it appeared that manipulations of structural complexity did not elicit an ERP (exclusively) related to syntactic working memory processes. It seems that retrieval mechanisms are partly dependent on prosodic and attentional processes and possibly semantic representations – the propositional content. Also, the manipulation of the linguistic structure to be retrieved appeared to impact the (proposed) integration phase. To further investigate retrieval and integration processes, an experiment was designed to focus on the semantic aspect. As I have pointed out in Chapter 2.4.2, quantifying expressions may be a burden on mechanisms of movement and/or copying since additional structural information has to be analysed. Extending this insight to processing, a Copy α account predicts a structural processing cost during recovery of ellipis in which quantification is contained. For example, a processing cost is expected in Stripping constructions when measured at "too" in (1b), when compared to (1a). If the recovery process is contingent of the retrieval of a full-fledged syntactic structure, this should be reflected as a syntax-related ERP.

(1) a. Mira peeled the oranges for breakfast, and Erica too.
b. Mira peeled all oranges for breakfast, and Erica too.

In so far as a pointer can be understood as direct reference to a proposition, a cue-based account predicts a relative ease of retrieval of a proposition that involves quantification. To the extent that such a small effect can be measured, it is expected to be reflected as part of the early positivity as found in the previous chapter. It is further expected that ERPs relating to the integration phase may be undetectable if the modulation of structure (and the inherently propositional content) is relatively small.

7.2 Methods

Test materials

From the data set described in Chapter 5.3, forty-two stimulus pairs were chosen from Stripping conditions *d* and *e*. These are repeated here in (2) as conditions ***a*** and ***b***). Condition *a* can be contrasted with condition *b* to estimate the difference between a determiner and a quantifier in Stripping constructions. As explained, items within such a pair maximally differed 1.25 in their average acceptability scores. The range of average scores among chosen items was 4.38-6.50 and the means of scores did not differ between conditions [$t(41) = 1.41, p = .166, d = .218$].

(2) a. Mira schilde de sinaasappels voor het ontbijt, en Erica **ook**.
Mira peeled the oranges for the breakfast and Erica too
'Mira peeled the oranges for breakfast,
and Erica too. ' (***Determiner Stripping***)

b. Mira schilde alle sinaasappels voor het ontbijt, en Erica **ook**.
Mira peeled all oranges for the breakfast and Erica too
'Mira peeled all oranges for breakfast,
and Erica too. ' (***Quantifier Stripping***)

From the remaining pretested item sets, 21 items from the first condition and 21 from the second condition of item set (8) in section 5.3 were chosen as fillers to prevent participants from expecting a certain type of ellipsis in the second conjunct. For a complete list of test sentences and fillers and average acceptability scores see Appendix B. Note that the measure point of interest in the test sentences is *ook*. If only sentences of the Stripping type in conditions ***a*** and ***b*** were used, the participants will start to predict at the point of the word *en* how the sentence will proceed – probably resolving the ellipsis before the moment *ook* appears. Therefore, Gapping conditions as shown in (3) were added as 'Related Fillers' (a duplet is used by means of an example).

(3) a. Koen verving de kast in de woonkamer, en Judith de lamp in
Koen replaced the cabinet in the living room and Judith the lamp in
de gang.
the hall
'Koen replaced the cabinet in the hall, and Judith the lamp in the hall.' (*Determiner Gapping*)

b. Koen verving elke kast in de woonkamer, en Judith de lamp
Koen replaced every cabinet in the living room and Judith the lamp
in de gang.
in the hall
'Koen replaced the cabinet in the hall, and Judith the lamp in the hall.' (*Quantifier Gapping*)

Filler sentences were selected in such way that the differences in rating means did not differ between the two filler sentence types (Determiner Gapping and Quantifier Gapping) as a paired t-test showed [$t(20) = 0.35, p = .73, d = .076$]. An additional set of 42 unrelated fillers were added. Between fillers and test conditions, the difference in means was kept as small as possible. A one-way ANOVA was run to establish the difference of the means between four groups (Determiner Stripping, Quantifier Stripping, Related Fillers and Unrelated Fillers) each consisting of 42 items [$F(3, 164) = 1.61, p = .19, \eta^2 = .028$]. Table 7.1 summarises the descriptive statistics of the selected stimuli. The two test conditions were divided over two lists and combined with the related and

unrelated fillers in such a way that only one item per test pair was presented once to each participant.

Condition	Mean	N	Std. Error
Determiner Stripping	5.46	42	0.10
Quantifier Stripping	5.32	42	0.08
Related filler	5.52	42	0.08
Unrelated filler	5.55	42	0.06
Total	5.46	168	0.04

Table 7.1: Means of rating of selected test sentences and fillers.

Participants

Twenty-two right-handed native Dutch participants with normal or corrected-to-normal vision took part in this study and were paid €15. The EEG data of one participant was not recorded due to technical failure. Three others were discarded from the analysis due to too many artefacts resulting in fewer than eight trials in one of the conditions. Of the remaining 18 participants six were male and the mean age was 22.28 (range 18-28). 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

Participants were comfortably seated in a dimly lit sound-proof room at a distance of approximately 90 cm of a 19 inch LCD monitor. One-hundred-and-eight test sentences were presented in a random order using the presentation software E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA). The sentences were presented word by word in Verdana font (36pt). Each word was presented for 300 ms with a 300 ms fixation cross interval. Presentation of a trial started with a fixation cross for 1,000 ms. Every sentence was followed by a content question to encourage comprehension. A blank screen with a duration of 1,500 ms intervened between the last word of a trial and the presentation of the comprehension question. For half the participants the left response button referred to "YES", for the other half the left button referred to "NO". Participants were given a break after 12 sentences and could proceed at their own pace. The comprehension questions referred to different parts of the sentences equally. Before the actual test, the participants were able to get used to the task with four practice sentences.

The experiment was concluded with a working memory test as used before.

In total, the experiment took about 1.5 hours per participant, 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, Bell, Jung, & Sejnowski, 1996). Remaining artefacts were rejected on the basis of the same criteria as used in all earlier experiments reported in this thesis. Epochs of 1,000 ms were computed with a 200 ms pre-stimulus baseline and ERP grand averages were time-locked to the onset of the target words *de* and *elke*. 14.15% of the trials were excluded from the analysis; of the 21 trials presented per condition, an average of 18.03 ($SE = 0.73$) were retained.

In the current experiment, a repeated measures ANOVA was planned using within-subjects factors CONDITION (two levels: Determiner Stripping, Quantifier Stripping), ANTERIORITY (3 levels: frontal, central, posterior), and, for analyses involving lateral sites, HEMISPHERE (2 levels: left, right).

7.3 Behavioural results

On average the accuracy on the comprehension questions of the two test conditions was 96.70% ($SE = 0.65\%$). The accuracy scores were similar for both test conditions [$M_{\text{Determiner}} = 96.83\%$, $M_{\text{Quantifier}} = 96.56\%$] as the difference in mean values was not significant as shown by a paired t-test on the scores [$t(17) = 0.19$, $p = .848$, $d = .046$]. The accuracy scores of the test conditions (Determiner Stripping and Quantifier Stripping) were further compared to the related filler conditions (Determiner Gapping and Quantifier Gapping). A significant difference was apparent as shown by a repeated measures by subjects on the scores [$F(3, 51) = 17.77$, $p < .001$, $\eta^2_G = .366$]. As can be seen in Table 7.2, which reports the results of a multiple comparisons procedure with Bonferroni correction, mean comprehension accuracy differs significantly between Stripping and Gapping conditions but not within these conditions.

			Comparison (p-value)			
Condition	Mean ratio	*SE*	**D Strip**	**Q Strip**	**D Gap**	**Q Gap**
D Strip	.968	.010	-	1	$<$ .001	$<$ .001
Q Strip	.966	.010		-	$<$ .001	.005
D Gap	.876	.016			-	1
Q Gap	.881	.019				-

Table 7.2: Means and standard errors of the accuracy scores and p-values of the Bonferroni-corrected multiple comparisons of the test conditions (**D**eterminer **Strip**ping and **Q**uantifier **Strip**ping) and related filler conditions (**D**eterminer **Gap**ping and **Q**uantifier **Gap**ping).

The accuracy of the three test sessions of the working memory task was on average 70.00% (SE = 2.79%). Per condition, the scores were: $M_{\text{Random Counting}}$ = 67.78% , $M_{\text{Auditory Presentation}}$ = 58.89%, $M_{\text{Visual Presentation}}$ = 83.33%. A repeated measures ANOVA by subjects with CONDITION as independent factor and ACCURACY OF NUMBER RECALL as dependent variable showed that the scores differed between conditions [$F(2, 34) = 7.75$, $p < .043$, $\eta^2_G = .202$]. A multiple comparison with Bonferroni correction showed that the visual condition differed marginally from the random condition [$p = .055$] and significantly from the auditory condition [$p < .001$].

The scores from the sentence comprehension task were compared with the scores from the working memory task. A large and significant correlation was found between the variables ACCURACY OF SENTENCE COMPREHENSION and ACCURACY OF NUMBER RECALL [$r = .632$, $p = .005$].

7.4 Electrophysiological results

Figure 7.1 depicts the grand averages of Determiner Stripping condition (***a***) and Quantifier Stripping condition (***b***) at the critical measure word *ook*. Most prominently, and mainly at frontal and central electrodes, a positive deflection starting around 300 ms after onset is apparent in both conditions. At some electrodes, condition ***b*** seems to deviate from condition ***a*** in a more positive direction. However, a permutation test per sample at every electrode with independent factor CONDITION did not yield significant time windows to be analysed any further.

Since the correlation between accuracy of sentence comprehension and the scores on the working memory task was significant in this experiment, its relevance was explored. First, the relation between working memory scores and ERPs were taken into account, and then, the relation between sentence comprehension and ERPs.

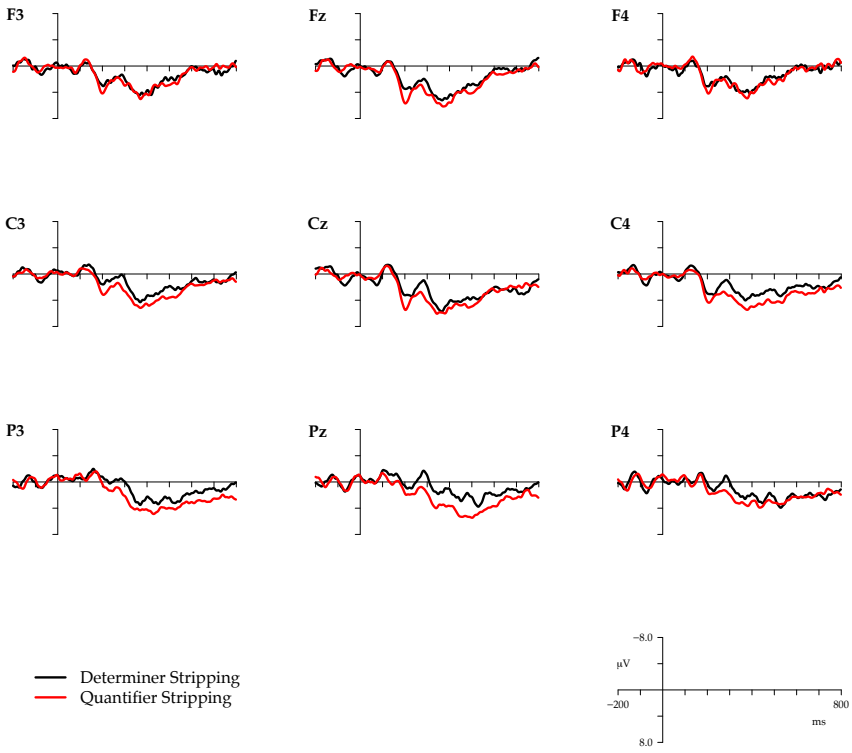

Figure 7.1: Grand averages of Determiner Stripping condition (***a***) and Quantifier Stripping condition (***b***) at onset (y-axis) *ook* at electrode sites F3, Fz, F4, C3, Cz, C4, P3, Pz and P4. Corresponding example sentences can be found on page 135.

7.4.1 Relation between working memory and ERPs

On the basis of a median split, the participants were divided into two groups, (*i*) consisting of participants with relatively low accuracy on the working memory task (range: 40.00%-66.67%, M = 57.08%, SE = 2.96%) and (*ii*) consisting of participants with relatively high scores (range: 73.33%-93.33%, M = 82.96, SE = 1.96%). Figure 7.2 depicts the grand averages of the first group while the second group is shown in 7.3. In each group, a permutation test per sample at every electrode with independent factor CONDITION was conducted. No effects were found. I also wished to explore the possibility of overall processing differences between groups, by taking the average across conditions and comparing it between the two groups. The result can be seen

in Figure 7.4. A permutation test per sample at every electrode with independent factor GROUP did not show any effect.

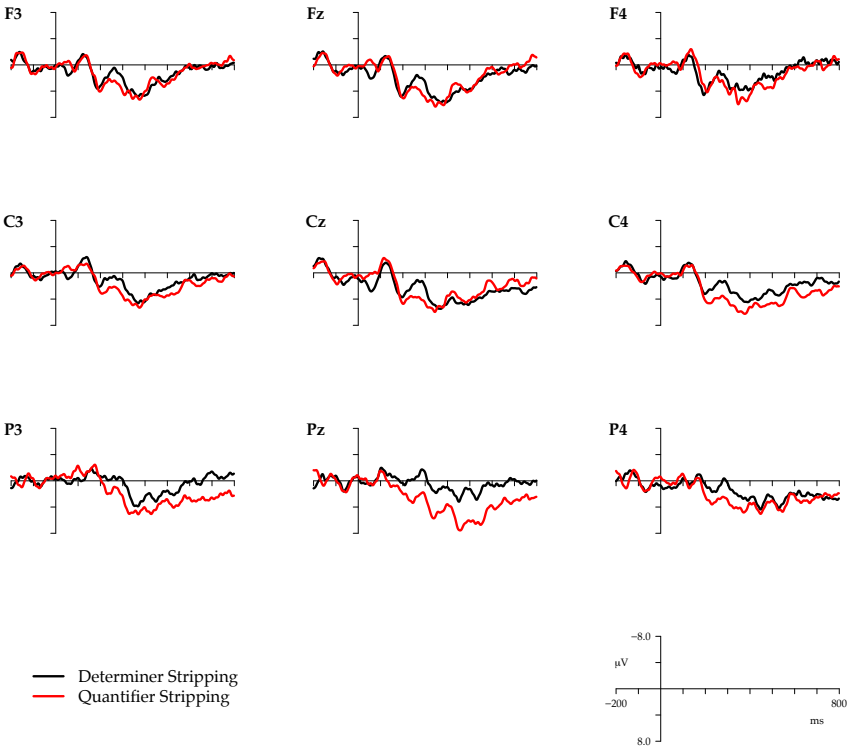

Figure 7.2: Grand averages of participants with low working memory scores (n = 9) of Determiner Stripping condition (***a***) and Quantifier Stripping condition (***b***) at onset (y-axis) *ook* at electrode sites F3, Fz, F4, C3, Cz, C4, P3, Pz and P4. Corresponding example sentences can be found on page 135.

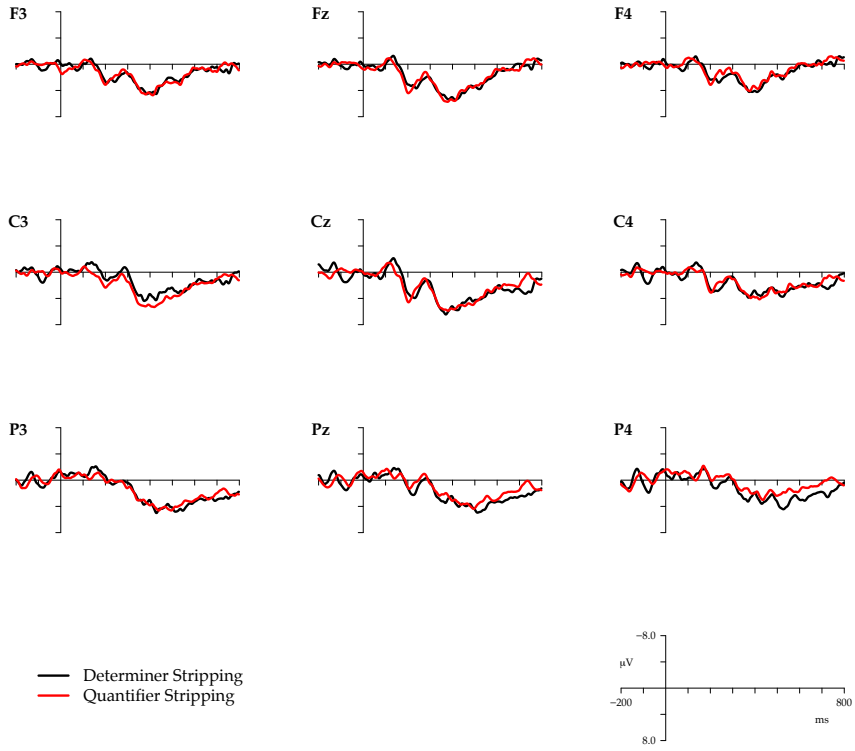

Figure 7.3: Grand averages of participants with high working memory scores (n = 9) of Determiner Stripping condition (***a***) and Quantifier Stripping condition (***b***) at onset (y-axis) *ook* at electrode sites F3, Fz, F4, C3, Cz, C4, P3, Pz and P4. Corresponding example sentences can be found on page 135.

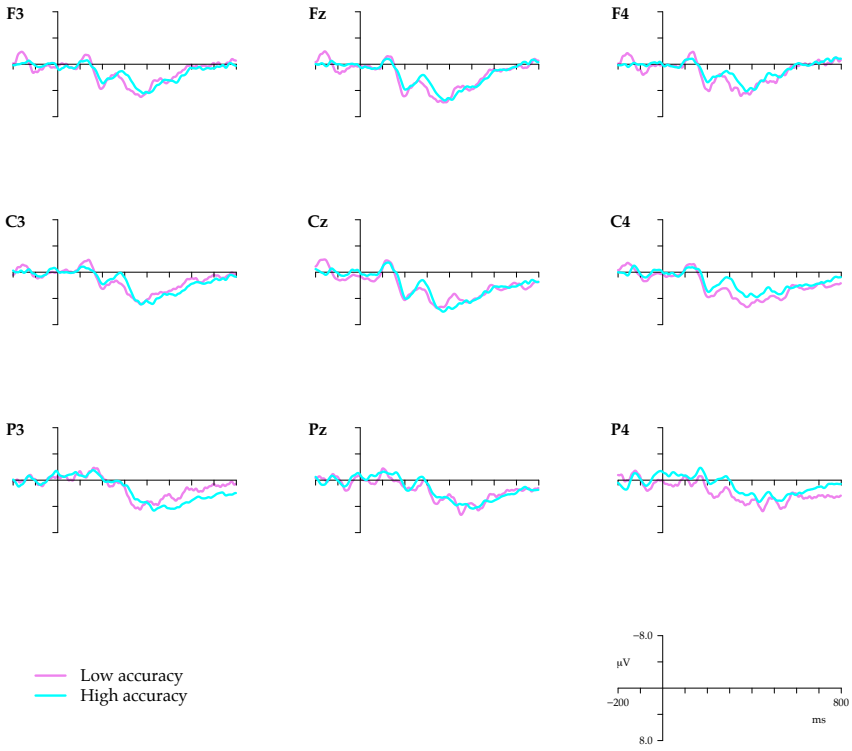

Figure 7.4: Collapsed grand averages of conditions ***a*** and ***b*** of two groups split by working memory scores: 'Low accuracy' (n = 9) and 'High accuracy' (n = 9) at onset (y-axis) *ook* at electrode sites F3, Fz, F4, C3, Cz, C4, P3, Pz and P4. Corresponding example sentences can be found on page 135.

7.4.2 Relation between sentence comprehension and ERPs

Again, on the basis of a median split, the participants were divided into two groups, (*i*) consisting of participants with relatively low accuracy on the sentence comprehension task (range: 82.54%-88.89%, $M = 87.48\%$, $SE = 0.67\%$) and (*ii*) consisting of participants with relatively high scores (range: 89.68%-96.83%, $M = 93.65$, $SE = .78\%$). Figure 7.5 depicts the grand averages of the first group while the second group is shown in 7.6. In each group, a permutation test per sample at every electrode with independent factor CONDITION was conducted. No effects were found. Also, the possibility of overall processing differences between groups was explored, by taking the average across conditions and comparing it between the two groups. The result can be seen in Figure 7.7.

Based on a secondary permutation test per sample at every electrode with independent factor GROUP, two repeated measures ANOVAs were run for a time window between 340-800 ms with within-subjects factors CONDITION, ANTERIORITY and HEMISPHERE (at lateral sites), and between-subjects factor GROUP. The means differed between groups significantly on midline sites [$F(1, 16) = 5.28$, $p = .004$, $\eta^2_G = .086$] and on lateral sites [$F(1, 16) = 7.75$, $p = .001$, $\eta^2_G = .164$]. Further, an effect of ANTERIORITY was apparent on lateral sites [$F(2, 32) = 11.31$, $p < .001$, $\eta^2_G = .127$]. A post hoc multiple comparison with Bonferroni correction showed that the means of amplitudes at central electrodes were equally significantly more positive than frontal and posterior electrodes [$p < .001$, $M_{\text{Central}} = 2.99$ ($SE = 0.33$), $M_{\text{Frontal}} = 1.29$ ($SE = 0.21$), $M_{\text{Posterior}} = 1.51$ ($SE = 0.25$)].

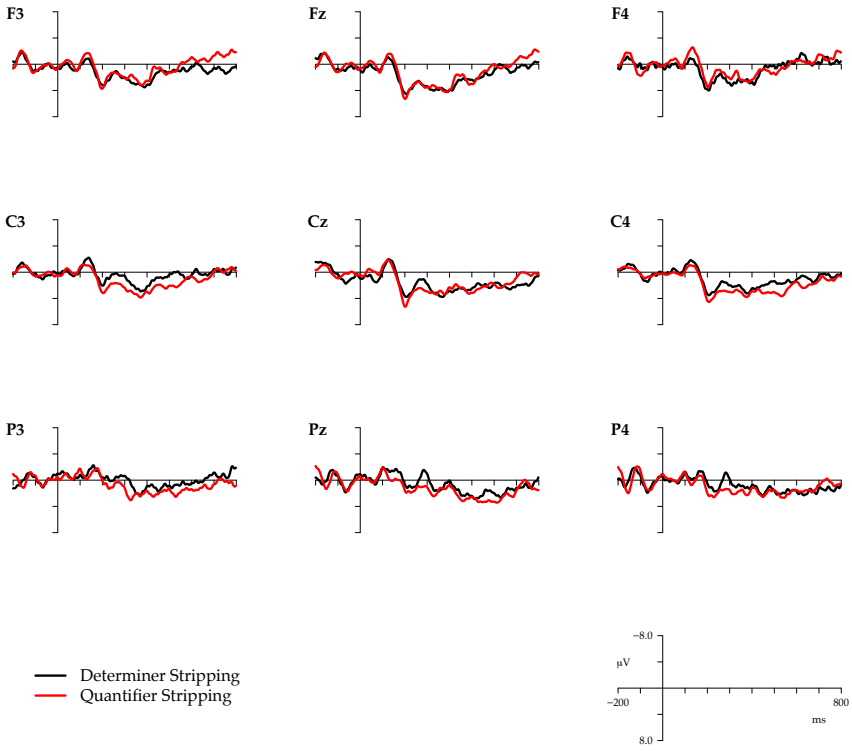

Figure 7.5: Grand averages of participants with low sentence comprehension scores (n = 9) of Determiner Stripping condition (***a***) and Quantifier Stripping condition (***b***) at onset (y-axis) *ook* at electrode sites F3, Fz, F4, C3, Cz, C4, P3, Pz and P4. Corresponding example sentences can be found on page 135.

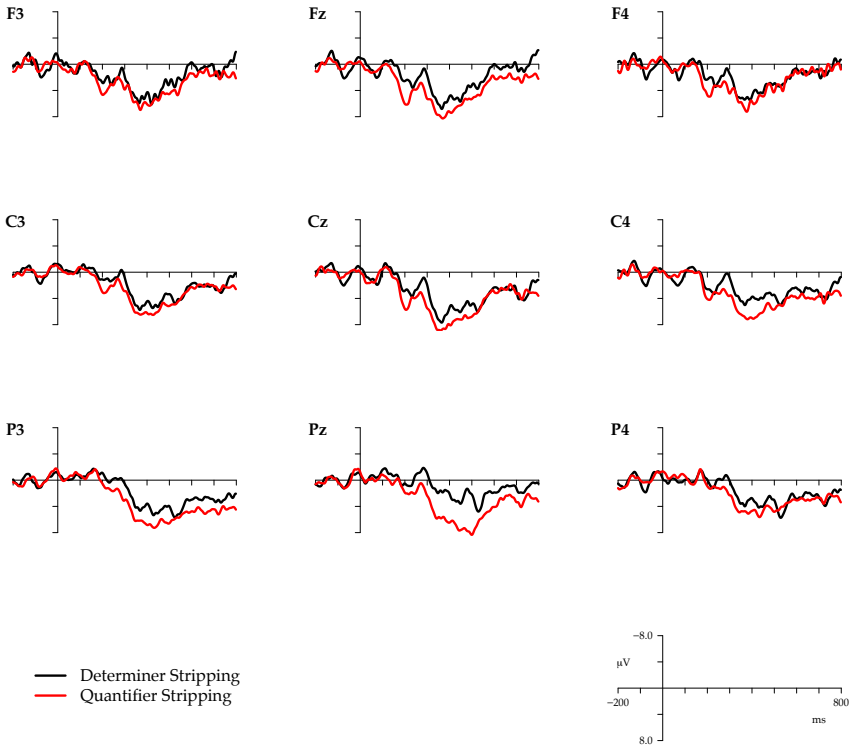

Figure 7.6: Grand averages of participants with high sentence comprehension scores (n = 9) of Determiner Stripping condition (***a***) and Quantifier Stripping condition (***b***) at onset (y-axis) *ook* at electrode sites F3, Fz, F4, C3, Cz, C4, P3, Pz and P4. Corresponding example sentences can be found on page 135.

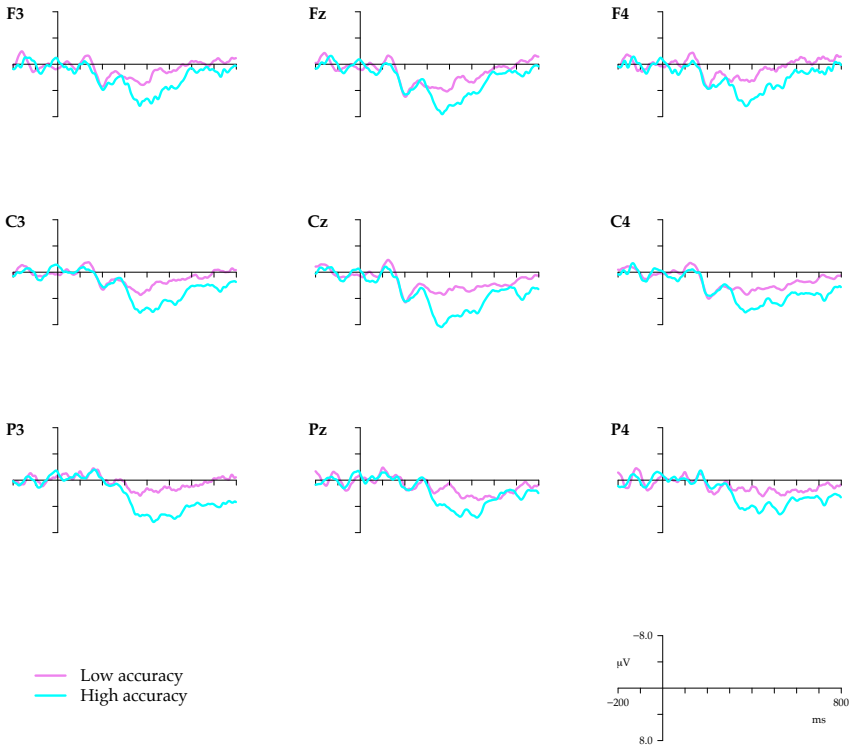

Figure 7.7: Collapsed grand averages of conditions ***a*** and ***b*** of two groups split by sentence comprehension scores: 'Low accuracy' (n = 9) and 'High accuracy' (n = 9) at onset (y-axis) *ook* at electrode sites F3, Fz, F4, C3, Cz, C4, P3, Pz and P4. Corresponding example sentences can be found on page 135.

7.5 Discussion

In this experiment, I investigated the difference between determiner *de* "the" and quantifiers *elke/alle* "every/all" in Stripping constructions. I hypothesised that if the recovery process is contingent on the retrieval of a full-fledged syntactic structure, this should be reflected as a syntax-related ERP. Contrastingly, I hypothesised that a pointer can be understood as direct reference to a proposition and that, as a consequence, a cue-based account should be able to account for a relative ease of retrieval of a proposition that involves quantification. Further, I expected that this retrieval process to be reflected as part of the early positivity as found in the previous chapter. I also expected that ERPs relating to the integration phase might be undetectable if the manipulation of structure (and the inherently propositional content) is relatively small.

The comprehension scores of the sentences in the main test conditions (Stripping) were at ceiling. Although rating means of these sentences did not differ from those of the related filler sentences (Gapping), the Stripping sentences were easier to comprehend. Note, that the rating means as collected in the pretest were based on the structure as well as the interpretability. While in the first experiment on structural modulations described in Chapter 6.1 no difference in comprehension scores was apparent between the Gapping conditions (***b***-***c***) and the Stripping condition (***d***), in the current experiment, it appeared that Gapping sentences were relatively more difficult to understand. Since in the current experiment rating means of the presented items were better matched than in the first experiment on structural modulations reported in Chapter 6.1, it seems reasonable to conclude that the current comprehension difference follows from the fact that in the Gapping conditions there are three contrasting phrases instead of one, thus, adding more information load to the utterance.

There was a tendency for positive deflections, starting around 300 ms after onset of the critical word in the grand averages of the whole group of participants in both conditions. At some electrodes, the positive deflection of the condition with the elided quantifier seemed larger as compared to the condition with the determiner. However, the positivity did not yield a significant difference. Meanwhile, a significant correlation was found between sentence comprehension scores and the working memory task scores (in contrast to the previous three ERP experiments reported in this thesis). Participants with high scores on the working memory task generally had high scores on the sentence comprehension task. I explored to what extent the working memory scores and sentence comprehension accuracy might be related to the ERP results.

While no effects could be established with respect to the relation between working memory scores and the ERPs found in the sentence reading task, it appeared that the positivity in both conditions as observed in Figure 7.1 was mainly generated by the group of participants with high scores on the sentence comprehension task (c.f. Figure 7.7). Considering the fact that in both

test conditions a recovery process of Stripping was involved, the conditions were collapsed and compared between groups. It appeared that the group with high accuracy on sentence comprehension showed a sustained positivity starting at 340 ms. Possibly, with a larger group of participants it may be demonstrated that this sustained positivity actually consists of two positive components as found in earlier experiments. If these results can be replicated with a larger group, the question is why there is a difference in processing strategies. As discussed earlier, the positivity may be related to several mechanisms, including processes related to attention. For the time being, I tentatively conclude that participants with relatively high comprehension accuracy were actively involved during the reading task while others were passive interpreters which may have caused slightly more difficulty during the comprehension task.

Aside from this interesting exploration, the rationale behind this experiment was to investigate to what extent quantifying expressions may help to decide between two mechanisms of ellipsis resolution. I hypothesised that quantifying expressions are a burden on a mechanism such as Copy α which would predict a structural processing cost during the recovery of the ellipsis. This should be reflected as a syntax-related ERP. However, the results seem to uphold the contrasting view that a possible antecedent for ellipsis is navigated in memory using cues which point to a more fully interpreted linguistic structure which may consist of additional information types. As a consequence, ERPs related to the integration of this structure can be measured. However, it appeared that the difference between determiners and quantifiers in stripped expressions is too small to the extent that neither the retrieval nor the integration phase are reflected by a distinct ERP. Still, additional experiments are needed to confirm this.

7.6 Conclusion

In this experiment, I used a semantic manipulation to study the processing of Stripping constructions. To do so, I compared the determiner *de* "the" with quantifiers *elke/alle* "every/all" in Stripping constructions, comparing ERPs at the critical measure point *ook*.

On the hypothesis that Copy α predicts a structural processing cost during the recovery of the quantified structures, a syntax-related (early) ERP was expected as a reflection of this mechanism. Although visually a slight difference was apparent in ERPs, the difference between the determiner and quantifier conditions was not significant. Additionally, the timing and polarity of the deviance was comparable with the early positive component as found in previous experiments, on the basis of which I concluded that retrieval processes are not exclusively steered by syntax-related mechanisms. It seems that this conclusion also applies to the current experiment.

Related filler sentences with Gapping constructions were presented and

all sentences were matched in terms of acceptability. As a consequence, a straightforward comparison could be made as to the comprehension scores of Stripping versus related Gapping conditions. This comparison indicated that Gapping sentences are relatively more difficult to understand than Stripping constructions.