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

Ruijgrok, B.J.

CHAPTER 3

Experimental background of Gapping and Stripping

In this chapter I explore the experimental literature on ellipsis. This line of research follows characteristic issues as raised by the theoretical literature to a certain extent. I take into consideration two behaviourally motivated parsing models that are grounded in theoretical insights and that are proposed as a possible link between theory and data. Again, I examine the role of prosody, as well as the relevant ERP components that have been found in relation to the recovery of elliptical structures.

3.1 Objectives and methods of experimental ellipsis research

Successful sentence comprehension requires the listener to parse syntactic, semantic and prosodic information. During parsing, the language processor encodes this information (processes it into memory) and probes it, for example, when earlier processed items need to be retrieved at a later moment. In reading sentences, the same types of information seem to be at play, including prosody – despite the fact that we typically read silently, sound coding appears to be an integral part of the reading process (Pollatsek, Treiman, & Pollatsek, 2015) with punctuation serving to convey prosody to a large extent (Rietveld and Heuven 2001:234; see also the "Implicit Prosody Hypothesis" by Fodor 2002). Assuming that a model for ellipsis resolution should account for the identification, reactivation, and integration of the antecedent, experimental research on ellipsis has entertained the following questions (see also Phillips & Parker, 2014):

- What is the time course of ellipsis resolution?
- To what extent do the antecedent and elided material need to have the same syntactic form?
- Is there syntactic structure at the ellipsis site at all?
- What is the effect of antecedent complexity?
- What is the effect of distance between antecedent and ellipsis site?

These questions may help to show to what extent parsing and interpretation of 'normal' sentences differs from the parsing of elliptical sentences. Different behavioural measures have been used to investigate ellipsis, such as acceptability judgements, comprehension questions, reaction times and reading times.

The questions listed above differ from theoretical questions to the extent that experimental linguists try to develop dynamic models with a focus on the timing of processes. By extension, experimental linguists utilise "judgements" of utterances in a different manner. During acceptability judgement tasks, respondents are for example, asked to specify how acceptable they find certain sentences. Acceptability is usually defined in terms of well-formedness and interpretability of an utterance. In the theoretical literature, argumentation hinges on the grammaticality of sentences. This literature tends to employ a categorical notion of grammaticality: usually, such sentences must be deemed either grammatical or ungrammatical, to sustain a particular theory. By and large, theoretical scholars gain information about grammaticality judgments by consulting their own intuitions. However, it appears that some sentences may have an in-between status, as shown by the examples in (1). Contrasting theoretical linguists, experimental linguists try to explain the gradual difference between these sentences in terms of the relative difficulty to process them which may relate to the time it takes to interpret the utterance.

(1) a. None of the astronomers saw the comet, but John did.
b. ?Seeing the comet was nearly impossible, but John did.
c. ??The comet was nearly impossible to see, but John did.
d. *The comet was nearly unseeable, but John did.

[Arregui, Clifton Jr., Frazier, and Moulton (2006)]

During an acceptability test, speakers are asked for their judgments on a range of sentences. To do so, respondents may be asked to use a rating, or "Likert" scale. Crucially, the way data are collected sets experimental research apart from non-experimental research. During experimentation a variable may be manipulated, for example, the grammaticality of sentence (1a). The manipulated grammaticality can be seen in (1b), (1c) and (1d). It is further required that every other factor that might influence the response is controlled as much as possible. This encompasses a wide variety of factors, ranging from the amount of words in the test sentences to the way participants are instructed, to name just a few. Furthermore, it does not suffice to invite twenty (or so) participants and show them a list of sentences. It may be, for example, important to intermingle experimental sentences with filler sentences if participants are required to remain oblivious to the effect that an experimenter is after. Usually, the sentences are presented in a random order.

Finally, acceptability judgements should be seen as only a first step towards experimental research, as it is not a measure that is generally accepted to be directly informative about underlying processes in the brain. Being an offline task, this method requires participants to think about sentences at a meta-linguistic level – comparable to the theorists approach. This is not the level we are primarily interested in if we want to know how the language system works. Therefore, as a technique it is best-suited for probing intuitions or to pretest stimuli for use in an experiment, that is, to use it as a tool to control variables. For example, for a certain task, experimental sentences should not differ too much in terms of understandability.

Employing a Likert scale in an acceptability judgment task allows the respondents to rate the stimuli on a five- or seven-point scale. Other distributions are possible, but the advantage of an uneven scale is that there is a middle point which may be interpreted as "no opinion" (of course, if a researcher wants to obtain a forced choice an even scale should be used). Further, since participants are likely to avoid the extremes, a seven-point scale may be preferred, if one expects different degrees of acceptability. As an alternative to traditional judgement procedures in the syntactic literature, Bard, Robertson, and Sorace (1996) proposed a Magnitude Estimation (ME) method, which is common in the psychophysics field. In an ME task, participants are asked to estimate the acceptability of a sentence by using their acceptability rating of a different sentence as a unit of measure. However, Sprouse (2011) shows that commutativity does not hold for acceptability judgements: only 20% of the participants were able to decide whether sentences that were equal in terms of difficulty were indeed comparable.

Provided that all stimuli are grammatical, comprehension questions can be used as a method to establish the relative difficulty of utterances. If the researcher is measuring, for example, EEG during reading, it is desired that the participants silently read the presented material. To ensure that participants actually perceive a presented stimulus sentence a comprehension question that targets the content of the stimulus can be presented after the reading task. Although offline tasks – such as acceptability judgements and comprehension questions – have been connected to the relative *processing* difficulty of sentences, in general online measures such as reaction time and reading time are regarded as appropriate behavioural measures to get insight in online structure computation (of whichever information type). Importantly, experimental designs that use online measures (including EEG) usually make reference to a theory (or 'model') that can account for temporal aspects. Before we take a look at psycholinguistic models that have been used to account for ellipsis data, we will briefly return to the discussion of surface versus deep ellipsis.

3.2 Surface versus deep

We have drawn from the theoretical literature that Gapping can be understood as a surface anaphor. This notion has also been advanced in the realm of psycholinguistic approaches. Such approaches have attempted to translate anaphoric relations into "recovery clues". Some research has focused on syntactic parallelism in deep- and surface-ellipsis to put the theoretical proposal of Sag and Hankamer (1984) to the test (e.g. Murphy, 1985a, 1985b; Tanenhaus & Carlson, 1990). A surface anaphor is assumed to bear a linguistic recovery clue by means of the remnant(s) (the sentence subject and often the object). Since a verbal context is required, the possible antecedent is a predicate relation. This relation is assumed to be suitable as a linguistic antecedent. Deep anaphors require recourse to a discourse model.

Murphy (1985b:792) suggests that "the motivation for the deep/surface distinction is related to difficulty in recovering the correct antecedent" and therefore he suggests it is a psychological distinction rather than a linguistic one. He argues that Gapping is part of surface anaphora since a linguistic antecedent, that is literal information about the antecedent, is necessary to interpret the ellipsis. In (2) the relation between subject and object needs to be recovered in order to interpret the clause "and Amy Carl".

(2) a. John duped Bill, and Amy Carl.
    b. John believed Bill, and Amy Carl.

[Murphy (1985b)]

In addition to the relation between the antecedent subject and object, "the exact form in which the relation was originally expressed" needs to be retrieved (Murphy, 1985b:803). Hence, the interpretation depends heavily on whether

this relation is encoded and remembered as "John duped Bill" or "John believed Bill". Since a structural representation is needed for a successful interpretation, at least for Murphy, this is a powerful reason why Gapping cannot be pragmatically controlled and hence should be regarded as a form of surface anaphor concluding that "conjunction reduction must have a linguistic antecedent by definition, regardless of its recoverability or processing characteristics" (Murphy, 1985b:806).

Note, that – complying with Sag and Hankamer (1984) – for Gapping, an LF representation should be appropriate. Murphy further endorses the notion of copying of syntactic structure, available in short term memory, as a mechanism to resolve surface anaphors. What is confusing is that a statement like "the copying process [...] must have access to the surface form of the antecedent" (Murphy, 1985a:296) may be in accordance with Hankamer and Sag (1976)'s proposal but not with Sag and Hankamer (1984) – given that the surface form of the antecedent is not an LF representation. On the whole, Murphy appears to argue for a parser that may consult syntactic and semantic information types; for him, overt syntactic structure is needed as well as a predicate relation, which might as well be represented in terms of the kind proposed by Dalrymple et al. (1991) and followers.

Using stimuli such as in (3), Tanenhaus and Carlson (1990) showed that, when anaphors are judged to make sense, structural parallelism has a facilitating effect on the speed of processing of both deep and surface ellipsis. This is unexpected if one follows Sag and Hankamer (1984) that would predict that deep anaphors such as "it" in (3c) should not be affected by structural parallelism.

(3) a. Someone had to take out the garbage. (*parallel to condition c and d*)
b. The garbage had to be taken out. (*non-parallel to condition c and d*)
c. But Bill refused to do it. (*deep anaphor*)
d. But Bill refused to. (*surface anaphor*)

[Tanenhaus and Carlson (1990)]

However, Tanenhaus and Carlson found a substantial effect on surface anaphors when no strict parallelism was available, with acceptability ratios being relatively low. In other words, surface anaphors made sense more often in syntactically parallel contexts than in non-parallel contexts. At the same time, parallelism did not affect judgments of deep anaphors. Since an interaction was found between syntactic parallelism and the type of anaphor, they interpreted this finding in favour of Sag and Hankamer. Thereby, Tanenhaus and Carlson sustained the claim that there exists a representational difference between surface and deep anaphors. Instead of a copy mechanism as assumed by Murphy, Tanenhaus and Carlson proposed that antecedents may be linked to an anaphor by means of a "pointer", noting that "copying is not the mechanism by which a surface anaphor would be associated with its antecedent in most current linguistic theories" (Tanenhaus & Carlson, 1990:261). Nonethe-

less, as we will see in the next section, the debate between copy and pointer approaches would persist.

3.3 Parsing strategies of ellipsis resolution

As mentioned in the first chapter, in the psycholinguistic literature a general distinction can be made between two approaches to sentence processing:

- syntax-first: serial
- constraint-based: parallel

The main motivation for an autonomous syntactic module as proposed by Frazier (for a comprehensive review see Frazier, 1987) was to reduce the burden on working memory load. She countered the assumption of parallel systems, where multiple syntactic analyses for an utterance may be computed and stored at the same time. Furthermore, she suggested that discourse-related factors do not influence an initial syntactic analysis. Her "garden path model" is sustained by abundant evidence that disambiguation of a sentence such as (4) – a famous example – is costly.

(4) The horse raced past the barn fell.

[Bever (1970)]

The syntactic analysis of *raced* is ambiguous since it could be the main verb, or it could be the beginning of a reduced relative clause (where *that was* has been omitted, i.e. *The horse that was raced...*). Parallel models would predict that both analyses compete with each other until the end of the sentence – keeping both sentence structures in memory. A serial model predicts that only one interpretation is being evaluated. By means of the principle known as "Minimal Attachment" (Frazier, 1979) – interpret a sentence in terms of the simplest syntactic structure – *raced* would be treated as main verb. If this interpretation crashed an alternative would be computed in turn. My impression of the serial-parallel discussion, however, is that the division is not clear-cut. Serial models may allow for implementing nonstructural factors during initial stages of syntactic analysis, therefore yielding interactive models. In other words, serial models may exploit different levels of representation in a parallel way. In serial computational approaches, this is most evident. For example, Lewis and Vasishth (2005)'s *Adaptive Character of Thought-Rational* (ACT-R) based serial parser would construct one fully formed analysis of (4) at a time while exploiting different information types. Based on activation decay and retrieval interference, their account utilises strategies that are mainly dependent on probabilities. It may be classified as an intermediate approach, between syntax-first and constraint-based. And, as we will see, it may be related to an ellipsis processing strategy.

A recent representative of a generative parallel constraint-based approach is the computational "surprisal" architecture of Levy (2008) in which "the parser allocates different amounts of resources to different interpretations of the partial input, and difficulty arises when those resources turn out to be inefficiently allocated" (2008:1128). In this model, partial input may be structural or lexical information, that is to say, incoming information is structurally analysed. Levy puts his architecture somewhat in between traditional serial and parallel models since it allows multiple structural variants to facilitate – rather than compete with – the processing of sentences such as (4). The relative difficulty of processing a word is understood as the word's surprisal given its context. In (4), *raced* is highly probable after encountering a noun phrase in English, which amounts to a preferred structure. The author suggests that his surprisal theory should be compatible with reading time accounts based on Lewis and Vasishth's model. However, it is as yet unclear how a surprisal account would handle elliptical structures.

Lewis and Vasishth's model implements a form of predictive parsing that has become known as "left corner parsing": a syntactic structure is built incrementally on a roughly word-by-word basis while little by little predictions may be made about the subsequent structure. Parsing is driven by a bottom-up as well as a top-down (predictive) mechanism. For example, such a parser for Dutch may predict a verb in the second conjunct of (5) as it assumes a structure that parallels the first conjunct.

(5) De man kocht een boek en de vrouw <*e*> een cd.
the man buy.3SG.PST a book and the woman <*e*> a cd
'The man bought a book and the woman a cd.'

Based on behavioural research, it has been suggested that in such sentences the antecedent structure will be activated even before the ellipsis has been detected (Callahan, Shapiro, & Love, 2010). Already at *en* in (5) the processor may use this conjunction as cue to expect a parallel structure upon which *kocht* will be reactivated. Callahan et al.'s evidence for the reactivation of previously processed antecedent information from memory is based on missing verbs and objects. However, it is not clear to what extent other additional phrases such as adjuncts will be reactivated, since they are possible candidates for ellipsis in Gapping and Stripping too.

An important issue concerns the *form* of the reactivated phrase. In accordance with a syntax-first approach, Frazier and Clifton (2001) propose that ellipsis may be resolved by inserting a copy of the missing structure (e.g. *kocht*). This predicts a low processing cost, regardless of the size of the antecedent. This mechanism, known as "Copy α" is assumed to be invoked when elliptical structures are encountered, substituting a step-by-step structure building procedure (for example, the default manner of parsing). Frazier and Clifton's evidence is drawn mainly from reading times that do not show any effect of complexity. Participants performed a self-paced reading task; in this task, they

were asked to read text frame-by-frame on a computer screen, pressing a button each time they were ready to move to the consecutive frame. Stimulus texts differed in complexity, as can be seen in (6) (forward slashes indicate the separation of the frames). Crucially, (6a) and (6b) differ in terms of the complexity of the antecedent: that is, in (6b), two clauses need to be retrieved as opposed to one in (6a). Nonetheless, in pairs such as (6), the reading times of the VP-ellipsis "Tina did too" did not differ between (6a) and (6b).

(6) a. Sarah / left her boyfriend last May. / Tina did too.
b. Sarah got up the courage / to leave her boyfriend last May. / Tina did too.

[Frazier and Clifton (2001, 2000)]

Copy α, however, may not be applicable for all ellipsis types. Notably, Gapping is regarded as an anomaly since it may allow for ambiguous syntactic scope. For example, sometimes the first noun phrase in a right conjunct can be interpreted as a secondary object of the first verb phrase, after which the interpretation crashes. This can be observed in (7), if *de vrouw* is interpreted as object of the verb *zag*.

(7) De man zag een boek en de vrouw <*e*> een cd.
the man see.3SG.PST the book and the woman <*e*> a cd
'The man saw a book and the woman a cd.'

Gapping thus prevents α from being straightforwardly determined and as a consequence, Gapping should be subject to a different parsing approach. This reminds us again of the surface versus deep discussion. Two different parsing strategies could be assumed for the deep-ellipsis in (8a) on the one hand and the surface ellipsis in (8b) on the other.

(8) a. Sarah bought a book. Tina did too. (VP-ellipsis)
b. Sarah bought a book. Tina too. (Stripping)

Just as Gapping, Stripping seems to allow for ambiguous readings changing the syntactic scope. "Tina" in (9b) may be agent or patient. Therefore, a copy mechanism would not be applicable to Stripping.

(9) a. Sarah broke up with her boyfriend last May. Tina did too. (VP-ellipsis)
b. Sarah broke up with her boyfriend last May. Tina too. (Stripping)

More recently Clifton. Jr. and Frazier (2010) have proposed that ellipsis (in general) may be constrained by structural and discourse conditions. Structural conditions would be provided by the grammar while the application of discourse conditions are assumed to be a quality of the processor. Maintaining a copy mechanism (in subsequent work this notion evolved into structure "sharing" (Frazier & Clifton, 2005) and "recycling" (Arregui et al., 2006;

Frazier, 2013) of syntactic structure, would mean that its applicability is determined by the grammar. At the same time, it is once again clear that a characterisation of ellipsis requires an interactive approach as the choice of the antecedent in (9b) requires access to different information types. Something to keep in mind in relation to copying or sharing structure is that memory for surface aspects (e.g. syntactic structure) appears to be relatively short-lived, in contrast to semantic information which can still be accessed after longer stretches of discourse. Garnham and Oakhill (1987:614) refer to this phenomenon as "one of the best-established results in the psycholinguistic literature.". This compares with the short-lived nature of the phonological code for a word, which becomes available when the word is accessed (Levelt, 1999) but decays within four to seven words (Baddeley, 2012; Tanenhaus, Carlson, & Seidenberg, 1985).

Others have argued that, in general, accessing a copy of some structure should take more time since a serial search must be undertaken. For example, Martin and McElree (2008) have proposed an alternative view of ellipsis resolution that involves step-by-step structure building, while a more fully interpreted discourse representation of the antecedent is accessed and integrated. Implemented as a memory-based content-addressable pointer mechanism, it would involve direct access of the antecedent, suggesting that the speed of interpreting the ellipsis does not depend on antecedent complexity. Taking Copy α as competing account, Martin and McElree (2008) propose that a pointer should not be limited to mere syntactic structure; rather it might as well be linked to a "fully interpreted discourse representation" (2008:883). Their approach not only bears resemblance to pro-form theories of ellipsis, but it is also closely connected to the ACT-R approach as proposed by Lewis and Vasishth (2005). Notably, Callahan et al. have suggested that some aspects of predictive parsing may be underpinned by such a cue-based method.

In one experiment, Martin and McElree asked participants to determine the acceptability of sentences as in (10) and (11) (choices: yes/no). Again, forward slashes signal frames of phrase presentation, but in this experiment the frames were presented at predefined moments (i.e. it was not a self-paced task). Participants were required to respond after every sentence.

(10) *Simple antecedent*

a. The history professor / understood Roman mythology, / but the principal / was displeased to learn that / the over-worked students / attending summer session / did not.

b. *The history professor / understood Roman mythology, / but the principal / was displeased to learn that / the overly worn books / used in summer session did not.

(11) *Complex antecedent*

a. The history professor / understood Rome's swift and brutal / destruction of Carthage, / but the principal knew the over-worked students / attending summer session / did not.

b. *The history professor / understood Rome's swift and brutal / destruction of Carthage, / but the principal knew the overly worn books / used in summer session / did not.

[Martin and McElree (2008)]

Martin and McElree found that participants' accuracy on the acceptability task and response time in condition (11) was comparable to condition (10). The authors hypothesised the response time to be a reflection of the time needed to retrieve an antecedent and to interpret it at the ellipsis site. Since the response time was not affected by the complexity of the antecedent, they used this finding as evidence for a pointer mechanism during which a pointer directly accesses an antecedent. In their comprehensive review of experimental approaches to ellipsis processing, Phillips and Parker (2014) note that to establish the acceptability of these test sentences, it suffices to link the subject of the embedded sentence ("the over-worked students" versus "the overly-worn books") to the head of the antecedent VP "understood" while the object content may be disregarded. They further question the statistcial power of (Frazier & Clifton, 2001)'s Copy α study, a problem that has been overcome in follow-up experiments reported in Frazier and Clifton (2005).

Additional support for Martin and McElree's account is based on non-effects of distance between antecedent and ellipsis. Tested sentences as in (12) showed that distant antecedents yielded poorer accuracy of acceptability judgement, but crucially, the processing speed was not compromised.

(12) *Near antecedent*

a. The editor / admired the author's writing, / but the critics / did not.

b. *The editor / admired the author's writing, / but the binding / did not.

(13) *Distant antecedent*

a. The editor / admired the author's writing, / but everyone / at the publishing house / was shocked to hear that / the critics / did not.

b. *The editor / admired the author's writing, / but everyone / at the publishing house / was shocked to hear that / the binding / did not.

[Martin and McElree (2008)]

What strikes me most is that both Copy α and the cue-based mechanism predict comparable results: no processing cost. At least for the cue-based method,

this prediction holds if no additional referents intervene between antecedent and ellipsis. "Interpreting additional material decreases the specificity of retrieval cues" which may amount to "*cue overload* where the cues that make contact with the correct constituent in memory are insufficient for successful retrieval" (Martin & McElree, 2011:330). All in all, this leaves us with inconclusive evidence for structural complexity effects in particular. In the present study, we will explore the possibility of differentiating between the proposed mechanisms by means of electrophysiological data.

3.4 Prosody

As we have seen in the discussion on Copy α, Gapping-like constructions may lead to ambiguous readings. A seminal behavioural study on the effects of parallelism on the interpretation of Gapping is Carlson (2001). Maybe unsurprisingly, an ambiguous reading is dependent on (*i*) the thematic role restrictions of the elided verb, that is, the possible object the elided verb selects; (*ii*) the prosodic contour; and (*iii*) the discourse context if available. Using a written questionnaire using sentences as in (14), Carlson found that participants favour a Gapping analysis when a parallel structure between arguments is available (encouraged by italicising them). The percentage of Gapping interpretations is indicated in brackets.

(14) a. *Alice* bakes cakes for tourists and *Caroline* for her family. (81%)
b. Alice bakes *cakes* for tourists and *brownies* for her family. (3%)
c. *Josh* visited the office during the vacation and *Sarah* during the week. (40%)
d. Josh visited *Marjorie* during the vacation and *Sarah* during the week. (4%)
e. *Dan* amazed the judges with his talent and *James* with his musicality. (21%)

[Carlson (2001)]

Clearly, italicising parallel phrases influence interpretation, but note that still a preferred reading of conjoined objects is apparent in condition *c*. In other words, prosody as imposed on written input is an additional factor. However, it cannot overrule an alternative thematic fit imposed by the verb.

In a follow-up experiment, Carlson tested the hypothesis of Minimal Attachment as discussed in the preceding section (on page 54). Based on this principle she derived the hypothesis that the simplest legitimate syntactic analysis of an input is preferred during parsing. She compared this to a parallelism constraint that dictates that the most parallel analysis of a conjoined structure is preferred, which holds that the parser should look for similar syntactic roles. Carlson used stimuli such as explained in (15) and presented them auditorily. Immediately after presentation of a sentence, a question appeared

on-screen asking the participants to choose the best paraphrase of the sentence they had heard.

(15) a. Josh visited the office during the vacation and Sarah during the week. (*baseline prosody*)
b. [BOB insulted the guests during DINNER] and [SAM during the DANCE]. (*cooperating Gapping prosody*)
c. Bob insulted [the GUESTS during DINNER] and [SAM during the DANCE]. (*cooperating Non-Gapping prosody*)

[Carlson (2001)]

A baseline prosody predicts that when the ambiguous DP "Sarah" contrasts equally with either the first conjunct argument ("Josh") or "the office", the prosody will be compatible with both the Gapping and Non-Gapping interpretations of the sentence. Contrastingly, cooperating prosody can bias the interpretation toward the Gapping or Non-Gapping analysis.

The Gapping response rate for condition *a* was 38% and is comparable to the 40% seen in the written study. Carlson concludes that a special Gapping prosody is not necessary for choosing a Gapping interpretation. The status of baseline was further corroborated by the Gapping response ratios for conditions *b* (44%) and *c* (28%), putting the baseline (almost) in-between. Again, a paralleled focused argument could not overrule a minimal attachment approach of the parser.

3.5 Electrophysiological research on ellipsis

As I have explained in Chapter 1.3.3 event related potentials (ERPs) can be measured to investigate the interplay between semantic, syntactic and (to a lesser extent) prosodic processes. During an ERP experiment, the EEG signal is analysed relative to specific time points in the experimental presentation. For example, when the time point at which a stimulus appears on-screen, researchers investigate how the brain activity responds to that particular stimulus. The average signal per condition per participant per electrode is averaged to get a "Grand Average" per condition (and electrode). The difference between the Grand Averages per condition is what is called event-related potential (ERP). Typically, in reading experiments, sentences are presented word-by-word. The onset of the presentation of a critical word is then taken as measure point for the onset of the ERP. Five main markers have been established in the literature: CPS, ELAN, LAN, N400 and P600 (see Table 1.1).

Given that the time resolution of electrophysiological methods is very precise, the focus in ERP research on ellipsis is on the time course of processing. Three steps in this time course may be considered:

- identification of missing structure
- reactivation of the antecedent
- integration of the retrieved antecedent

These steps are derived from two ERP studies on Gapping. The first study, carried out by Kaan, Wijnen, and Swaab (2004), investigated the time course of identification and resolution of verb gaps in English. Using sentences as in (16) they manipulated the plausibility of critical noun phrases ("the hammer") that followed a verb gap. For convenience, the elided verb is indicated between <>.

(16) a. Ron took the planks, and Bill <took> the hammer.
b. Ron sanded the planks, and Bill <sanded> the hammer.

[Kaan et al. (2004)]

Sentences were presented word-by-word and the task of the participants was to click either 'GOOD" or 'BAD" in order to rate the stimuli as a means of acceptability task.

The authors hypothesised that the determiner of the critical NP would be the first possible point for the processor to detect a missing structure – a verb. Since they recorded a centro-posterior negativity they suggested this to be a variety of an ELAN effect which may be connected to phrase structure violations. In addition to the early negativity, a positive fronto-central deflection between 300-500 ms was measured, which they cautiously ascribe to the retrieval of the preceding verb information.

What is problematic is that Kaan et al. compared the critical determiner to any determiner appearing in any sentence in this study. This leads to the appearance of unexpected effects, since the determiners that were used as baseline appeared in different positions in a sentence. As such, the established negativity does not make a strong argument for the detection of missing structure. Despite this, the authors propose a general mechanism of syntactic persistence reminiscent of the Copy α routine that extends beyond ellipsis: "the grammatical structure of a sentence can be stored in the working memory as an autonomous entity, and re-accessed in subsequent processing" (2004:590).

Kaan et al. reasoned further that if a missing verb were reactivated (they also refer to this as "reconstruction" of missing structure, here), integration difficulty would be apparent at the critical noun. With respect to this difficulty an N400 was expected, and was found. This was interpreted as evidence for reconstruction of the antecedent at "hammer" or just after detection of the gap (at the determiner). Note that an N400 cannot be interpreted in terms of the syntactic form of the ellipsis construction, but in terms of the lexical representation of the antecedent and possibly the relative difficulty of integrating this antecedent. In addition to their expected N400 effect, a P600 was found. They suggested that, possibly, the semantic anomaly (e.g. "sanded the hammer") may have induced a process of syntactic revision. This seems odd since there

is no difference in terms of syntax between plausible and implausible reconstructed phrases, that is, the anomalous verbs do not yield a different syntactic construction. Since a P600 may be larger when violations are task-relevant (see for a discussion Sassenhagen, Schlesewsky, & Bornkessel-Schlesewsky, 2014), it may be triggered by an acceptability task (as was used by Kaan et al.). Possibly due to meta-linguistic processes that are at work during an aceptability task, the authors can not be sure that the P600 was a reflection of a combined syntactic and semantic integration difficulty.

Kaan, Overfelt, Tromp, and Wijnen (2013) is the only ERP study in which Dutch sentences containing Gapping are compared with "No Gapping" counterparts. Just as the earlier study on English described above, only a verb was elided. They focused on the moment when the ellipsis site in Gapping constructions is detected and the moment when the antecedent is accessed and integrated. However, in contrast to the English study, they investigated to what extent an ellipsis site can be anticipated, rather than focusing on the question of whether there is syntactic structure at the ellipsis site. The move towards anticipation rather than trying to establish some kind of structure coincides with the current fashion of probabilistic methods (as put forward by for example Hale, 2011; Levy, 2008). But also, it appears to be very difficult to demonstrate that there is structure at an ellipsis site, despite the wide range of experimental techniques that have been used in attempting to do so (see for an overview and discussion Phillips & Parker, 2014). In Kaan et al. (2013), the left anterior negativity found in Kaan's earlier study is reinterpreted in terms of expectations. If the parser does not anticipate an elided verb, this may result in processing difficulty. However, in the case that ellipsis is expected, no such ERP component should be found. An example set of the stimuli used in this experiment is shown in (17). Stimuli were presented on-screen word by word and participants were again asked to determine after each sentence whether it was good or bad. In (17), we see four conditions of which (17a) and (17c) are of the most interest: these allow us to compare plausible Gapping with plausible No Gapping sentences while the critical measure point is the NP "de bloemen". However, the status of the second clauses differs. In condition *a*, we see two conjoined main clauses, whereas in condition *c*, a main clause is followed by a subordinate clause.

(17) a. Anouk zond de kaart aan haar vader, en Julia <zond> de
Anouk sent the card to her father, and Julia <sent> the
bloemen aan haar moeder.
flowers to her mother.
'Anouk sent the card to her father, and Julia the flowers to her mother.' (*Plausible Gapping*)

b. Anouk schreef de kaart aan haar vader, en Julia <schreef> de
Anouk wrote the card to her father, and Julia <wrote> the
bloemen aan haar moeder
flowers to her mother.
'Anouk wrote the card to her father, and Julia the flowers to her mother.' (*Implausible Gapping*)

c. Anouk zond de kaart aan haar vader, terwijl Julia de bloemen
Anouk sent the card to her father, while Julia the flowers
aan haar moeder stuurde.
to her mother shipped.
'Anouk sent the card to her father, while Julia shipped the flowers to her mother.' (*Plausible control for a*)

d. Anouk schreef de kaart aan haar vader, terwijl Julia de
Anouk wrote the card to her father, while Julia the
bloemen aan haar moeder stuurde.
flowers to her mother shipped.
'Anouk wrote the card to her father, while Julia shipped the flowers to her mother.' (*Control for b*)

[Kaan et al. (2013)]

The authors argue that in a subordinate clause (as in condition *c*) verb Gapping is prohibited, and crucially, no verb Gapping is expected after "Julia" since a reader (or listener) may predict the location and appearance of a verb as soon as the conjunction has been processed. Therefore, "terwijl" indicates that the verb will be appearing at the end of the clause, while "en" introduces a main clause requiring a verb right after the first phrase. According to Kaan et al., conditions *c* and *d* therefore are proper control conditions as no verb is expected before the NP "de bloemen". The plausible versus implausible conditions were added to investigate when the elided verb is semantically integrated – similar to the study in English discussed above.

The critical measure points were the determiner and the noun in the phrase *de bloemen* 'the flowers'. To analyse effects on the determiner the grand averages of conditions *a* and *b* were collapsed and compared with collapsed conditions *c* and *d*. No effects could be found. However, only for participants who performed poorly on the task, i.e. participants who had problems determining

that a sentence such as (17b) was bad, ERP effects at the determiner showed a negativity between 400-600 ms after onset. The authors suggest that this negativity is a later instance of the ELAN found in their earlier study. They regard it as a LAN that has been suggested as being an index of prediction strength, as is proposed by Lau, Stroud, Plesch, and Phillips (2006).

Similar to the English study, a P600 effect was found in the implausible Gapping condition at the noun. Notably, the P600 was also apparent in plausible Gapping constructions. However, the N400 was only slightly visible on the Cz electrode and there was no significant effect here. Concluding that the integration of the elided verb is a relatively late process (i.e. 600 ms after presentation of the noun), the authors suggest that the involvement of a resolution mechanism for Gapping is similar to integrating a *wh*-phrase object with its verb as in, for example, "Which book did you buy?"(Kaan, Harris, Gibson, & Holcomb, 2000). Upon encountering the verb "buy", the earlier processed object "which book" can be integrated. Note though that this example differs qualitatively from ellipsis processing, since a *wh*-phrase object always awaits obligatory integration, in contrast to antecedents of ellipsis. Kaan et al.'s findings seem compatible with the integration part of Brouwer and Hoeks (2013)'s "Retrieval-Integration" account that I introduced in Chapter 1.3.3. It is, however, still unclear what the form of the retrieved antecedent might be.

Kaan et al. assumed that an absence of a LAN would be evidence of a top-down approach during which the parser already reactivates antecedent information at the connective *en* ("and") since it expects an ellipsis. This would contrast with a bottom-up approach which would infer an ellipsis site upon encountering missing structure. A LAN would then be a sign of gap detection. Some readers may take issue in general with the principle of "absence of evidence is evidence of absence". In other words, absence of a LAN may well mean absence of a bottom-up approach but this should be taken to entail the existence of a top-down approach. Furthermore, it is striking that the authors overlook the fact that Gapping is an optional process. That is, there could be a verb following *Julia*, one that contrasts with the verb in the first conjunct. What is expected then by the parser, is verbal information, but not verb Gapping exclusively. In addition, the object can be elided as well, leaving the parser uncertain as to how much information should be reactivated.

As shown in (17), a comma was used to make explicit a separation between the two clauses. Although participants were encouraged to interpret a sentence such as (17a) as a parallel coordination of two clauses, it might be that some participants applied a process reflecting minimal attachment yielding an object coordination of "aan haar vader en Julia". Note that, although such a parsing strategy was certainly not possible in all experimental stimuli, it could have affected the results. For example, we could interpret the observed LAN as a reflection of the parser resetting the "Minimal Attachment" principle. By and large, the LAN cannot be straightforwardly linked to any expectations the parser might have, let alone for plausible, say, grammatical instances of, Gapping.

In her doctoral thesis, Dimitrova (2012) reports two auditory ERP experiments on Gapping using prosody as modulation to disambiguate structurally ambiguous constructions. Using sentences as in (18) she measured ERPs at the subject of the second conjunct. Sentences were presented in isolation in one experiment and presented with a preceding question (that biased towards the applicable prosody of the following sentence) in another. "Good thematic fit" means that a minimal attachment applies, that is, a conjunction of two clauses is dispreferred.

(18) a. John invited PETER on Monday and MARTIN on Tuesday. (*preferred-good thematic fit-no Gapping prosody*)
Discourse: When did John invite the boys?
b. JOHN invited Peter on Monday and MARTIN on Tuesday. (*dispreferred-good thematic fit-Gapping prosody*)
Discourse: When did the boys invite Peter?
c. JOHN peeled the orange with a knife and MARTIN with his hands. (*enforced-poor thematic fit-Gapping prosody*)
Discourse: How did the farmers peel the orange?
d. John peeled the ORANGE with his knife and MARTIN with his hands. (*anomalous-poor thematic fit-no Gapping prosody*)
Discourse: What did John peel?

[Dimitrova (2012)]

In the trials where no preceding question was present to provide disambiguating discourse context, Dimitrova found in condition (18b) a marginal right-lateralised negativity (400-700 ms), followed by a positivity (700-1,000 ms) relative to (18a). In sentences with poor thematic fit (18c)-(18d), Gapping prosody elicited a broadly distributed negativity (400-700 ms). Nongapping prosody in (18d) triggered an anterior-central negativity (700-1,000) and a posterior positivity (700-1,300 ms) as compared to (18c).

When discourse contextualising questions did precede the sentences, a centro-posterior negativity (400-700 ms) was elicited irrespective of thematic fit in Gapping readings ((18b) and (18c)). In good thematic fit sentences, the Gapping bias as established by the preceding question yielded an anterior positivity and posterior negativity in two subsequent time windows: 700-1,000 and 1,000-1,300 ms.

Dimitrova suggests that the interpretation of an accented ambiguous element such as "MARTIN" in a Gapping reading yields an N400-like component – apparent in both experiments. She considers this component "to be related to the activation of verb phrase information and the assignment of a subject role (rather than an object role) to the accented element" (2012:228). She found the N400 in sentences with and without a biasing context. Just as we have seen in Carlson's study (2001), a Gapping reading for a sentence like (18b) is dispreferred. Dimitrova relates the established P600 to the reconstruction of a more complex (and dispreferred) structure, yielding two conjoined clauses.

However interesting these results are, it seems odd to attribute the negativity found in (18c), as compared to (18d), to the same underlying procedure as apparent in condition (18b). "MARTIN" in (18c) is compared with an anomalous condition. In other words, we cannot be entirely sure that the effect found was due to Gapping prosody alone; rather, they could be interpreted in the opposite direction – as is usual in experiments using anomalous conditions. To my knowledge, there is only one additional ERP (reading) study on ellipsis – although not on Gapping. In favour of the cue-based mechanism, the study of Martin, Nieuwland, and Carreiras (2012) shows that retrieval interference (possibly due to "cue overload") is reflected by a negativity between 400 and 1,000 ms after stimulus onset. Note, though, that this study concerns Noun Phrase ellipsis. This is crucially different from Gapping since no predicate relation is involved. Spanish sentences such as in (19) were presented to participants word by word. Some sentences (60% of the trials) were followed by a comprehension question. Between brackets, gender of a noun is indicated. The NP-ellipsis "another" has to match with the correct gender of the antecedent. In all sentences, there is an intervening noun (an attractor) that is structurally unavailable as antecedent.

(19) a. Marta se compró la camiseta que estaba
Marta REFL buy.3SG.PST DET.F t-shirt REL be.3SG.PST
al lado de la falda y Miren cogió
PREP+DET.M next PREP DET.F skirt and Miren take.3SG.PST
otra para salir de fiesta.
another.F to go.3INF PREP party.
'Marta bought the t-shirt (fem.) that was next to the skirt (fem.) and Miren took another (fem.) to go to the party.' (*correct attractor-same*)

b. Marta se compró la camiseta que estaba
Marta REFL buy.3SG.PST DET.F t-shirt REL be.3SG.PST
al lado del vestido y Miren cogió otra
PREP+DET.M next PREP DET.F skirt.F and Miren take.3SG.PST
para salir de fiesta.
another.F to go.3INF PREP party.
'Marta bought the t-shirt (fem.) that was next to the dress (masc.) and Miren took another (fem.) to go to the party.' (*correct attractor-different*)

c. *Marta se compró la camiseta que estaba
Marta REFL buy.3SG.PST DET.F t-shirt REL be.3SG.PST
al lado de la falda y Miren cogió
PREP+DET.M next PREP DET.F skirt and Miren take.3SG.PST
otro para salir de fiesta.
another.M to go.3INF PREP party.
'Marta bought the t-shirt (fem.) that was next to the skirt (fem.) and Miren took another (masc.) to go to the party.' (*incorrect attractor-same*)

d. *Marta se compró la camiseta que estaba
Marta REFL buy.3SG.PST DET.F t-shirt REL be.3SG.PST
al lado del vestido y Miren cogió otro
PREP+DET.M next PREP DET.F skirt.F and Miren take.3SG.PST
para salir de fiesta.
another.M to go.3INF PREP party.
'Marta bought the t-shirt (fem.) that was next to the dress (masc.) and Miren took another (masc.) to go to the party.' (*incorrect attractor-different*)

The increased negativity found in (19b) at the NP-ellipsis is considered as a retrieval interference. In the ungrammatical sentences (19c) and (19d), a sustained negativity was found. The authors concluded that "structurally unavailable noun phrases are at least temporarily considered for grammatically

correct ellipsis" (2012:1859). This would mean that cues are stored and activated regardless of the syntactic structure they originated from.

3.6 Summary and conclusions

In this chapter, I have highlighted the results of previous experiments related to structural complexity and prosody with a focus on Gapping. The experimental literature on ellipsis very much reflects the characteristic issues raised by the theoretical literature. Grounded on theoretical insights, models have been proposed as a link between theory and data. Two of these, Copy α and the cue-based mechanism, reflect to some extent the divide between syntax-first and constraint-based approaches. Although it is tempting to use these models to estimate the extent of syntactic structure available at the ellipsis site, I would like to quote Phillips and Parker (2014:15)'s conclusion that "caution is required in mapping findings about timing of ellipsis resolution onto theories of the representation of ellipsis constructions" .

Carlson (2001)'s experiments emphasise the role of prosody during Gapping resolution. Although apparent, prosody does not outweigh the influence of verbal information. It would therefore be interesting to see how different prosodic contours modulate conjunctions that are not ambiguous between Gapping and Non-Gapping. Just as in Dimitrova (2012), an ERP experiment could be designed to do so. As an alternative explanation of differences imposed by different prosodic contours, a lack of parallel intonation could be interpreted by the parser as a cue that the unaccented argument may be considered as possible structure to elide. Prosody, then, would help make predictions about upcoming structure.

Kaan et al. (2004)'s finding of ELAN in combination with a positivity between 300-500 ms at the determiner may be a reflection of a retrieval process, but we cannot be sure about the form of the retrieved material in this study. Although the authors suggest that the antecedent is reconstructed at that point as part of a general mechanism of syntactic persistence, retrieved information might be of another information type which is integrated once the object is processed. This integration process was tested more adequately in their follow-up study, comparing Gapping and similar Non-Gapping constructions. The P600 found in that study could well reflect an integration process. However, it remains unclear on which information type(s) of the antecedent this process was operating. For example, instead of dealing with fully-fledged structure, the integration phase may be confronted with a more fully-interpreted chunk.

It might be difficult to ascertain either information type in the ellipsis site. However, if we can tease apart the different predictions of the models discussed above, we might end up with a indication of how to map neuronal activity to representations proposed by the theoretical literature. In the next chapter, I will argue that this might be possible.