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

Ruijgrok, B.J.

Author: Ruijgrok, Bobby
Title: Tapping into semantic recovery : an event-related potential study on the processing of gapping and stripping
Date: 2018-05-31

CHAPTER 4

Setting the stage

In this chapter I discuss the mapping between existing theoretical insights and actual processing. I arrive at a comparison of Copy α and the cue-based mechanism with respect to the timing of processes of retrieval and integration. Since individual differences may lead to differences in (amplitudes of) ERP components and may be ascribed to natural variability in the capacity of human working memory, I further propose a suitable working memory test. I conclude with hypotheses and possible results.

4.1 Bridging theoretical and experimental research

4.1.1 Introduction

Linguistic research at the Leiden University Centre for Linguistics, the place where the current study was carried out, may be broadly described as the study of structure and variation among the world's languages. At this institute three types of linguists ranging from theoretical and descriptive to experimental can be found in different workspaces: the so-called armchair, the field, and the lab. All working towards an understanding of human language, it seems at times that their insights are difficult to reconcile. This chapter[1] describes a framework for those linguists who view language ultimately as a cognitive system.

The following analogy may be used to show that the division of workspaces need not lead to a segregation into distinct linguistic fields per se. Imagine Anne, sitting in her garden chair, noticing that the ants in her garden are walking faster as the temperature increases. While sitting there, she comes up with a function rule for this phenomenon. Her neighbour Eddy embarks on a jungle trek in South America and tries to apply the function to the Amazonian ants, without any success. However, he does notice that there is a high level of humidity. He decides to build a database in which he lists facts about temperature, humidity and walking speed of the different ants he finds. Already questioning the domain and range of Anne's function,[2] their mutual friend Onno checks the limits of Anne's proposed function in his beloved botanical garden, manipulating both temperature and humidity as possible factors. He further relates his findings to the physical properties of several kinds of ants. Finally, the three friends arrive at an integrated theory of the ant's walking speed.

While the link between abstraction and observation in the analogy is pretty straightforward, it is clear that three methods of investigation have all contributed to the understanding of the ant's behaviour. At the same time, predictions stipulated by their *shared* theory can easily be tested in different research domains. The idea is that the complementary aspect benefits empirical research, which encourages linguists in the armchair, field and lab to better understand each other, given that their research goal is the same. In what follows, I will show that different methods of data collection and different levels of analysis need not divide linguistics into separate fields.

[1]Section 4.1 has been published in Reckman, Cheng, Hijzelendoorn, and Sybesma (2017) and are slightly edited for this dissertation.

[2]This function appeared in an actual math assignment in the 1980s in the method *Getal en Ruimte* (Noordhof Uitgevers). To the amusement of the class, an ant could end up walking backwards at some degree below zero.

4.1.2 What is at stake?

Within a generative approach, the grammar system is usually assumed to be a static entity of knowledge that resides in the brain and that interacts with a processing system containing comprehension and production mechanisms. Ever since Chomsky's seminal work *Aspects of the Theory of Syntax* (Chomsky, 1965), the division between competence (grammar system) and performance (processing system) has fuelled linguistic analysis. However, 50 years later it also appears that the hypothesised division has constituted an obstacle for linguists who aim to link linguistic theory to neuro- and psycho-metrical data. This chapter will consider two apparent issues. Firstly, unlike the properties of physical phenomena such as the walking speed of an ant, we (still) lack a device that can objectively and directly measure the properties of a cognitive phenomenon such as "grammar". Although in the past decades a division of data collection has usually been linked with two separate language systems, we will see in section 4.1.3 that while such a division may be ideal in terms of theory, it is obscure in practice. The second problem concerns the linking of two separate systems. The question here is, how should the interaction between a grammar and separate processing system be defined? A possible (beginning of a) solution of this problem will be the topic of section 4.1.4.

4.1.3 Methods of measurement

Offline versus online data collection

Traditionally, linguistic theory finds its basis in categorical distinctions conceived of and assessed by means of introspective judgements – either those provided by the linguist, or informally collected by asking colleagues at work, conferences or other meetings. The use of judgement data that is collected by means of controlled experiments (either through web-based tools, in the field or in the lab) has been embraced by some, but it is still frowned upon by others (see for a discussion the special issue of *Theoretical Linguistics, 33* (3), 2007). A hypothesised split between introspection and experimentally collected judgements seems, however, untenable: although it has usually been taken as fact that introspection is a reflection of linguistic competence, operating beyond any kind of performance, one could say that this method of research forms one end of an empirical continuum, ranging from well-informed individual offline judgements gathered from colleagues, to online measures of a naive group of people taken in highly controlled experiments. In addition, corpus data collected in the field (i.e. systematic collections of naturally occurring texts of both spoken and written language) may be used to quantify linguistic phenomena, which may be useful at any level of the empirical continuum. For example, an experimenter may need to extract data from a corpus to check for possible confounds in a stimulus set. Whichever method is used, the commonality between linguistic researchers is that they try to generalise their results to the speech community.

During an offline judgement task, a participant (or a single linguist) responds to a certain linguistic stimulus with no time restrictions. Meanwhile, online measures such as reaction time, eye movements, brain potentials – to name a few – may give insight into online language structure computation. While a considerable part of linguistic theory is unconcerned with online processes, it is desirable that, if a theory purports to have computational strength, it is at least able to specify how computations are implemented during online language use. In this sense, experiments that measure reaction time, for example, may add to a (computational) theory of language just as judgement data do. However, it is unwise to use reaction time data to compare two theoretical constructs that both lack hypotheses about timing to start with.

On the assumption that offline responses reflect the representations of the grammar and online responses reflect processing mechanisms, Lewis and Phillips (2015) note that frequent misalignments between offline and online responses should be apparent. By "alignment" they mean the extent to which constraints of language processing are the same as those imposed by the grammar. Take, for instance, a garden-path sentence such as (1).

(1) Colleagues sent the invitation to Crit's retirement party were happy.

Only if the reader or listener is given enough time, an initial computation, in which *colleagues* is subject of *sent*, can be revised. Online the sentence may be judged ungrammatical, contrary to an offline response.

While it is indeed the case that misalignments exist, the authors effectively show that specific types of misalignment between online and offline responses amount to specific stages of computation. Crucially, the misalignments seem predictable. The authors therefore claim that:

> [...] online and offline representations are the product of a single structure-building system (the grammar) that is embedded in a general cognitive architecture, and misalignments between online ("fast") and offline ("slow") responses reflect the ways in which linguistic computations can fail to reflect the ideal performance of that system. (Lewis & Phillips, 2015:39)

Treating different methods of data collection as lying on a continuum naturally corresponds to Lewis and Phillips's view that representations of one language system can be investigated using different measures. As different methods target the same representations, the object of study is one system. The advantage of apprehending a single system is twofold. Firstly, it opens the door to incorporating gradient patterns that have been reported in both theoretical (usually indicated by question marks) and experimental (yielded by measurement type) research. Secondly, no separate account is necessary as to how those representations are identified during comprehension and how they are assembled during production. Future research is needed to confirm that a single system can carry out both comprehension and production tasks. In the remainder of this chapter I will explore which levels of description we need in order to define a cognitively-motivated language system such as this.

Online data in ellipsis research

Elliptical sentences, such as the Gapping example in (2), are interpreted by a process of "retrieving" and "integrating" earlier mentioned information (here: *bought a book* denoted by <*e*>).

(2) Eva bought a book in the shop, and Agnes <*e*> in the supermarket.

Interestingly, we can arrive at an interpretation of this sentence within a linguistic context without immediately available linguistic form. Though ellipsis is a multidimensional phenomenon since syntactic, semantic and prosodic constraints apply, theoretical approaches usually take one of these dimensions as a starting point to account for the nature and "recoverability" of the antecedent. For example, syntactic accounts generally represent the elided content as a fully-fledged structure at some point during the derivation, while semantic accounts would recognise the ellipsis as a more fully interpreted representation. A further issue concerns "licensing" which relates to the question of when ellipsis is allowed: which elliptical structures are well-formed? Although hybrid theories exist, an integrative theoretical account which incorporates syntactic, semantic and prosodic constraints of even one type of ellipsis is still to be developed, let alone a unified account of the phenomenon as a whole. Notably, Cremers (1993) argued that the interpretation of coordinate structures – including the ellipsis type as seen in (2) – is in part "extragrammatical", linking to a processing component.

In the psycholinguistic literature on ellipsis comprehension, it has been suggested that acceptability of ellipsis may depend on the amount of repair that is required to resolve omitted structure that does not exactly match the antecedent structure (Arregui et al., 2006). For example, using sentences such as (3) (repeated from Chapter 3.1) an acceptability decline ("gradience" if you will) can be observed, (3a) being judged most acceptable, and (3d) least acceptable. The example shows that a decline correlates with the relative difficulty the processor experiences in recovering the phrase *see the comet* in the right conjunct. The percentages of acceptable responses are between brackets.

(3) a. None of the astronomers saw the comet, but John did. (83%)
b. Seeing the comet was nearly impossible, but John did. (66%)
c. The comet was nearly impossible to see, but John did. (44%)
d. The comet was nearly unseeable, but John did. (17%)

[After Arregui et al. (2006)]

Although an independent grammar should abort any interpretation of ill-formed ellipsis constructions, it seems that they can be saved online. Constituting the ultimate tension between competence and performance, the question is why and when during comprehension the "parser" would overrule licensing instructions imposed by the grammar so easily, which in itself seems a licensing issue. Apparently, to develop linking hypotheses in a two-system approach one would need to account for mutually constraining factors. Within a single cognitive system that maps linear strings – sometimes incomplete (2) or sometimes (relatively) ill-formed (3) – to conceptual representations and vice versa, the attested gradience could be plausibly captured. Such a system would build representations of a grammar that amount to instructions ranging from higher-level (grammar) to lower-level (processing) procedures, theory being a kind of abstraction or idealisation of the parser (Sprouse & Almeida, 2013).

With respect to representations of antecedents of well-formed ellipsis constructions, experimental research has put forward at least two mechanisms that seem to fall on different sides of the familiar syntactic-semantic divide. Either a copy of *bought a book* in (2) (proposed as "Copy α" by Frazier & Clifton, 2001) or a more fully interpreted discourse representation that is directly accessible is inserted in the ellipsis site (implemented as a cue-based pointer mechanism by Martin & McElree, 2008). Both accounts assume that sentence comprehension is an incremental process during which incoming structural information is paired with an interpretation – updating representations step by step. In terms of retrieving and integrating the antecedent, both accounts predict, rather unhelpfully, the same behavioural results, stating that the speed of interpreting the ellipsis does not depend on antecedent complexity. For example, no difference would be observed between (2), repeated here in (4a), and (4b).

(4) a. Eva bought a book in the shop, and Agnes <*e*> in the supermarket.
b. Eva bought a book about gardening in the shop, and Agnes <*e*> in the supermarket.

However, it is important to distinguish retrieval from integration as it seems reasonable that a copy account may be rather beneficial to an integration process, while the relative cost of searching and finding structure might increase as a function of its size. Contrastingly, a cue-based account, which is mainly explaining the mechanism of retrieval, may predict the reverse. Figure 4.1 may help to explain how both mechanisms predict an *equal* processing cost with respect to the resolution process as a *whole*.

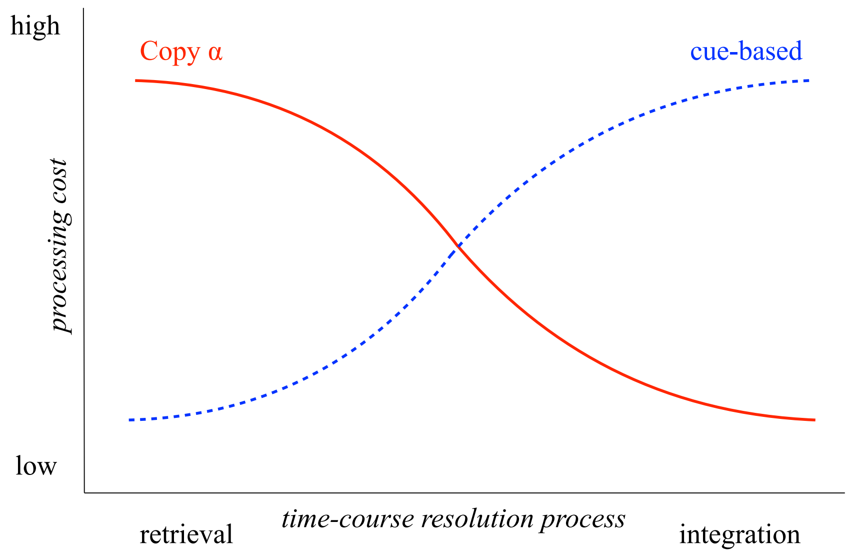

Figure 4.1: Schematic representation of predictions made by Copy α and a cue-based pointer mechanism, in terms of processing cost related to retrieval and integration processes.

To decide between the two mechanisms, the electrophysiological technique of event-related potentials (ERPs) may provide key insight, being the method of choice to investigate the time-course of cognitive processes. Effects on mechanisms of retrieval are expected early on, followed by those that impact on the integration process. In accordance with a copy mechanism, the onset of ERP signatures relating to accessing and copying missing structure would vary as a function of structure size; upon retrieval a fully-fledged structure would facilitate the integration process predicting relatively small effects. As mentioned, a "cue-based" approach would account for the reverse situation. Early ERP signatures of retrieval may be fixed since the antecedent is directly accessible. On the other hand, integration processes may operate on representations of various types, as discourse information has to be integrated in incrementally built-up structure, predicting ERP variability relatively late in the time course. Thus, ERPs may be used to compare models that are able to make predictions regarding timing. With this method, we may gain valuable insight with respect to the division of labour of syntactic, semantic and prosodic constraints. In the current study, these dimensions come under investigation – taking up the challenge of integrating theoretical conceptions of ellipsis resolution with cognitive performance data.

4.1.4 Towards a unified research program

Three levels of analysis

In the preceding sections, we have seen that a model based on a sharp distinction between static knowledge and processing mechanisms is not suitable to accommodate both theoretical and experimental research. One attempt to provide a framework that relaxes the competence-performance opposition has been put forward by Jackendoff (2002). Because this model emphasises the independent combinatorial character of syntactic, semantic and phonological information types, it seems to be particularly suitable for investigating the multidimensional character of phenomena such as ellipsis. However, his proposal is not sufficiently specific with respect to neurophysiological data such as ERPs to be truly integrative. The lack of a proper integrative theory has led (Poeppel & Embick, 2005:103) to provocatively forecast "(long-term) interdisciplinary cross-sterilisation rather than cross-fertilisation between linguistics and neurobiology, or, for that matter, linguistics and other empirical disciplines." In other words, we need a methodological framework that also incorporates physiological data. Furthermore, such a framework should specify hypotheses concerning the linking of theory and data.

Recently, Marr (1982)'s model for investigating vision has been put forward as reference to bring together linguistics and neuroscience (see for example Baggio, Lambalgen, & Hagoort, 2012; Embick & Poeppel, 2015). This model is built on three levels of analysis: (a) Computational Theory, (b) Representation and Algorithm and (c) Hardware Implementation. According to Marr any machine carrying out an information-processing task (i.e. a cognitive process) must be understood by answering the questions (a) what is computed?, (b) how is computation carried out? and (c) how can the computation be realised physically? An ideal integrated theory of language would then be a combination of formal theories of grammar, language processing and neural computation, respectively. Entertaining these levels as descriptions of one cognitive system, grammar could be understood as the abstract description of the representations that this system builds (Lewis & Phillips, 2015); representations that are identified and put together during comprehension and production, respectively. At the same time, such an integrated theory would help us to investigate explanatory connections between all three levels of analysis; for example, we could ask to what extent discoveries about the structure and functional organisation of the brain explain (rather than just describe) properties of the computations and representations that constitute language (Embick & Poeppel, 2015).

On the surface, the proposed framework may still resemble a competence-performance distinction, though with an added neurobiological level. It should be noted, though, that this system requires a theory of computation that can actually be carried out in real time. In that sense, "mentalistic" linguistic theory ought to proceed by according the same value to experimental results as it does to evidence from native speaker intuitions (Poeppel & Embick, 2005). Whereas in the "old distinction" a variety of phenomena were lumped together under the umbrella of performance, (Baggio et al., 2012:339) note that in the proposed approach these phenomena may be disentangled at a representational level and "understood in their distinctive features". For example, algorithms that specify working memory constraints may shed light on the type of data structures that a computational theory should produce, as well as the memory architecture and its neuronal substrates. Concentrating mainly on semantics, Baggio et al. stress – in line with Marr – the importance of the computational nature of integrating constraints derived from all levels of analysis.

Computational (psycho)linguistics

Hypothesis testing leads to theoretical progress. We have arrived at a model in which multiple sources of data can be taken into account and which enables us to test hypotheses locally, within levels. Preferably, they survive across levels. The less local the test domain, the more variables will have to be taken into account. The proposed integrated endeavour promotes a computational approach requiring rigid specificity which is simultaneously well-suited for testing hypotheses in highly controlled experiments. In terms of a division of computational approaches suggested by Cremers and Hijzelendoorn (2014), "gnostic", "paragnostic" and "agnostic" methods virtually align with the levels of description that we are now familiar with: formal theory, production and comprehension mechanisms and neural behaviour, ranging from "knowledgeable" (gnostic) explanatory linguistics to "naive" (agnostic) connectionist approaches in neurolinguistics. Intuitively, an integrated approach should aim at a computational model that combines symbolic and sub-symbolic terms, bridging the continuum of data collection methods and ultimately being explanatory at all levels of analysis.

In the meantime (computational) research proceeds step by step. This way, we will be able to determine the relative gnostic weight. (Lewis & Phillips, 2015:30) point to promising computational accounts that are based on transparent grammar-to-parser mappings, arguing that such models "may be understood as relating different levels of analysis, as in a one-system approach, rather than relating independent cognitive systems." The same transparency can be found in Cremers and Hijzelendoorn's ongoing project, "Delilah" (see for example Hijzelendoorn & Cremers, 2009; Reckman, 2009). Delilah is an example of a pure gnostic machine which parses and generates Dutch sentences on the basis of precise syntactic and semantic symbolic representations. Cremers and Hijzelendoorn (2014) also note the inevitability of incompleteness of grammar. The question is to what extent it can be supplemented by other terms than just symbolic ones. A "semantic machine" such as Delilah would require a semantic database – provided by computationally-based corpora research – other than the "lexikon", which is not (yet) available.

An example of a hybrid model of sentence processing is based on a well-established cognitive architecture *Adaptive Character of Thought-Rational* (ACT-R), proposed and primarily developed by John Robert Anderson (see act-r.psy.cmu.edu for a list of relevant applications and publications). Lewis and Vasishth (2005) developed a model that is able to simulate human reading time data. Utilising principles of memory retrieval and controlled processing, it is far from complete as its functionality only revolves around cue-based retrieval during syntactic parsing in the course of reading. However, it is flexible to the extent that it allows the researcher to add assumptions and theories about a specific task to be modelled. Furthermore, although a precise theory of cues is still lacking, an ACT-R based model may provide us with a tool to determine the nature of effects of interference, locality, antilocality and storage effects in sentence processing. It may turn out, to the dismay of some formal linguists, that some linguistic phenomena are grounded in principles of general cognitive processes. Yet, a means to estimate the limits of formal conditions is exactly what we need – even if one would still subscribe to the "old distinction". If we embrace a computationally sound approach, we can speak of one system that provides the representations that both listeners and speakers arrive at during language use.

4.1.5 Conclusion

I have argued that we should understand the human language system in terms of three levels of description that ultimately amount to a computational model. While I endorse the hypothesis that contradictory outcomes from offline versus online data may be due to different stages of the computations they tap into, future research is needed to confirm this. Computational linguistics may add valuable insights just as theoretical, descriptive and (other) experimental research does; furthermore, it has the benefit that it may provide us with data produced by highly controlled experiments. Testing computational models and integrating and manipulating the amount and type of predefined constraints will enable us to bridge theory and data – provided that real-time computation is a shared level of explanation. Within an integrated approach, we may determine which linguistic constraints are essentially linguistic and which of them are manifestations of more general cognitive capacities. A platform is at hand to overcome the persistent gnostic-agnostic divide.

4.2 Working memory load

Over the last few decades, the period in which neurobiological research has expanded dramatically, it has become clear that no human brain is the same. Although the gross (functional) anatomy seems to be similar among the population to some degree, there are certainly individual differences. If we assume that every human employs automatic linguistic processes, we may not expect to see differences in this area. However, in the field of language comprehension individual differences have indeed been found (see for example Kaan et al., 2013; Otten & Van Berkum, 2009). Such variation may lead to differences in (amplitudes of) ERP components and may be ascribed to variation of the capacity of people's working memory systems. This is expected to play a major role during ellipsis processing. To control for such variability, I will test participants by means of an additional memory task. This section explains the nature of this task and why we have chosen to use it.

4.2.1 A model of working memory

The term 'working memory' (WM) stems from the earlier proposed notion of short-term memory (STM). Following Baddeley (see for a history and overview Baddeley, 2012), I will regard STM as a system for "simple temporary storage of information, in contrast to WM, which implies a combination of storage and manipulation." It is understood by Baddeley as a multicomponential capacity comprising four subsystems:

- a phonological loop, concerned with verbal and acoustic information
- a visuospatial sketchpad, the visual equivalent of the phonological loop
- an episodic buffer, a multi-dimensional buffer store that links between WM components, but also links WM to perception and long term memory
- a central executive system, an attentionally-limited system which links to the episodic buffer

WM, as a whole, serves the function of integrating the information types that are processed by the phonological loop and visuospatial sketchpad into a unified representation. This representation may be stored for a short while and be manipulated upon. With regard to language comprehension, the episodic buffer is of particular interest: it has been suggested that its capacity may predict the aptitude of prose comprehension, as we will see below.

4.2.2 Working memory and sentence comprehension

The seminal study by Daneman and Carpenter (1980) suggests that there is a correlation between, in their terms, "WM span" and the capability for prose comprehension. Subsequent research, using paradigms where participants were required to employ a combination of temporary storage and processing, corroborated Daneman and Carpenter's findings (see for a meta-analysis Daneman & Merikle, 1996). Daneman and Carpenter (1980)'s paradigm – in which participants read out a series of sentences of different lengths while having to remember the last word in each sentence – has since become classic.

It has further been suggested that a separate subsystem of the WM is employed to assign syntactic structure to a sentence and to use this structure to determine the meaning of that sentence (e.g. Caplan & Waters, 1999). Caplan and Waters suggest that WM capacity as a whole might not be associated with differences in the efficiency of syntactic processing in sentence comprehension. Rather, the process of "recognizing words and appreciating their meanings and syntactic features; constructing syntactic and prosodic representations; and assigning thematic roles, focus, and other aspects of propositional and discourse-level semantics" (Caplan & Waters, 1999:78) might call on a different pool of resources. They propose a separate sentence interpretation resource theory, which assumes that general WM tasks cannot be used to predict language processing efficiency. We could interpret this standpoint as referring to the hypothesis that automatic linguistic processes are independent of general WM. However, this study was written before the construct of the *episodic buffer* had been put forward. This module was proposed to account for the fact that an executive system should be able to link to a temporary storage. Daneman and Carpenter (1980) and follow-up research had shown that such storage should be bigger than the limited capacities of the phonological loop and visuospatial sketchpad. The function of the buffer should be to integrate and maintain information into coherent episodes. Further, a proper link could now be established between WM and long-term memory. In sum, the episodic buffer is assumed to play a major role in "binding information from diverse sources into unified chunks" (Baddeley, 2007:148). In a commentary on Caplan and Waters (1999), Kane, Conway, and Engle (1999:102) note that: "[...] working memory capacity is needed only under attention-demanding circumstances, and, insofar as syntactic processing appears to be immune to divided-attention conditions, it likely occurs relatively automatically." While syntactic aspects of language comprehension may be carried out automatically, the episodic buffer may enable us to explain how relatively more demanding tasks in which information needs to be stored (such as in ellipsis resolution) are executed.

From a neurophysiological perspective, some authors have proposed that some capacity of the human brain might be specifically devoted to syntactic working memory which appears to be a "bilateral network of inferior frontal and superior temporal brain regions, with a left lateralisation within the inferior portion of the pars opercularis of the left inferior frontal gyrus (Brodmann Area 44) " (Fiebach, Schlesewsky, Lohmann, Cramon, & Friederici, 2005; Fiebach, Schlesewsky, & Friederici, 2001). These brain sites, then, could be understood as the neural basis for the subsystem as proposed by Caplan and Waters (1999). Tasks that require maintenance of information and more computation on that information (i.e. using the episodic buffer) have consistently activated the mid-dorsolateral frontal lobe, that is Brodmann Areas 46 and 9 (see for example Petrides, Alivisatos, Meyer, & Evans, 1993). The mid-dorsolateral region is believed to keep track of our thoughts and memories; indeed, poor maintenance and manipulation of information is associated with impaired dorsolateral regions (e.g. Cannon et al., 2005).

4.2.3 Testing working memory

Verbal WM capacity can be tested by means of reading span (as introduced above). In the spirit of Daneman and Carpenter (1980), Noort, Bosch, Haverkort, and Hugdahl (2008) designed a reading span test that is compatible across four languages. In this computer-administered test, five trials of 20 sentences are presented. However, note that this task might be too demanding in an experiment where participants have already carried out a sentence reading or listening task beforehand; therefore I will avoid this situation in my own experiments. Besides, span tests are regarded by Caplan and Waters (1999) as calling on another resource than syntactic processing. It has been shown that the mid-dorsolateral region is implicated in the monitoring and manipulation of information in working memory (Petrides, 2000). Ideally, I should use a design in which a stimulus needs to be temporarily stored and be recalled after intervening structure has been processed. Therefore I am looking for a method that assesses WM storage and processing in a relatively short time. Petrides et al. (1993) may offer such a method, which can be understood as a variant of Daneman and Carpenter (1980). Since Petrides et al. and his colleagues were using Positron Emission Tomography as measure (a procedure during which participants are injected with a short-lived radioactive substance), they wanted to minimise the scanning period. Hence, the design was compact. Below, I reproduce their description of the testing procedure in which participants carried out "self-ordered" and "externally ordered" number generation tasks .

> [...] the subjects were scanned with PET for 60 sec under three different conditions of testing. In the control condition, the subjects were required to count aloud from 1 to 10 at the rate of approximately one digit per second. They were told that when they reached the number 10, they were once again to start counting from 1 to 10 and continue in this manner until told to stop. In the self-ordered condition, the subjects were asked to say aloud, in a random order, the numbers from 1 to 10. They were asked to monitor carefully the numbers they gave so as not to repeat the same number more than once until all 10 numbers were reported. At that point they were to begin a new trial (i.e., a sequence), again generating numbers randomly from 1 to 10. The subjects were asked to start always from the number 1, because this would permit the experimenter, who was recording the responses, to know when a new trial had begun. As in the control condition, the subjects were told to generate the numbers at the rate of approximately one per second. An average of 5.25 trials (range, 4.5-6.0) was completed during scanning, with an average error of 0.9. An error was defined as a repetition or an omission of a number in a trial. In the externally ordered condition, the subjects were told that, during scanning, the experimenter would read out in a random sequence the numbers from 1

> to 10, omitting one of these numbers. The subjects had to monitor carefully the numbers read by the experimenter because, on completion, they would have to say the number that had been omitted. The experimenter would then administer another trial - i.e., read another random sequence of the numbers 1 to 10, again omitting one number that the subject would be required to report. The numbers were read out at the rate of approximately one digit per second. An average of 5.6 (range, 5.0-6.0) trials was completed during scanning and the subjects made an average error of 0.2 per trial.
>
> Before each scanning condition, the experimenter explained the requirements of the task to be performed and the subjects practiced the task once. The subjects kept their eyes open during scanning, but visual stimulation was reduced by dimming the lights within the scanning room and by surrounding the subject with black curtains.
>
> [Petrides et al. (1993:880)]

The advantages of this paradigm are that the performance of the participants can be related to a fixed control condition and the critical conditions are related to the mid-dorsolateral region. Furthermore, it is a task in which participants have to fill a gap in some sequence, a procedure that is at least in part reminiscent of the resolution of ellipsis in language. In addition, it takes at most ten minutes to complete.

4.3 Hypotheses and possible results

We have arrived at a comparison of two behaviourally motivated models of ellipsis processing that reflect the syntax-semantics divide in the theoretical literature to a certain extent, namely Copy α and a cue-based pointer mechanism. Despite this, it has also become clear that a mapping between existing theoretical insights and neurophysiological processes may not always be straightforward or even justifiable. Therefore, although it is my intention to integrate theoretical approaches with the investigation of processes at a neural level, the results of the current study should nonetheless be interpreted with great caution when trying to relate the findings to theoretical notions.

Each of the experiments in this study aims to investigate the online time course of the processing of Gapping and Stripping and to what extent it may be modulated by syntactic, semantic and prosodic factors. In Chapters 6, 7, and 8, I report ERP studies testing syntactic, semantic and prosodic variables respectively. Let us consider overall hypotheses based on the theoretical discussions presented in the foregoing chapters.

Effects of the manipulated variables on mechanisms of retrieval are expected early on, followed by those that impact on the integration process. A copy account predicts modulation of ERP signatures related to syntactic processes early in the time course, possibly manifesting as (E)LAN effects. The relative cost of searching and finding structure might increase as a function of the structure's size. Because this account proposes that a full structure is available, integration processes would be carried out with relative ease. Contrastingly, a cue-based account, which is mainly explaining the mechanism of retrieval, would predict the reverse. This account, would predict a burden on integration processes. Early ERP signatures of retrieval may be fixed since the antecedent is directly accessible. On the other hand, integration processes may operate on representations of various types, as discourse information has to be integrated in incrementally built-up structure, predicting ERP variability relatively late in the time course. Therefore, modulation of cues would presumably be reflected by a modulation of P600 effects. A caveat is in order here, as it is not always clear what exactly can be considered to be cues. I take it that a cue can be related to any information type that is stored in a more fully interpreted chunk: syntactic, semantic and prosodic. Since a cue is directly accessible during processes of retrieval, it is expected that ERP signals related to retrieval are relatively small. In addition, I hypothesise that the processor may exploit a composite of different cues, if needed. I have no prediction, however, as to the nature of a possible ERP signature (or signatures) that might be implicated therein. As discussed, there is not much ERP research on Gapping, so this study must be regarded as somewhat exploratory. While this limits the way I can 'stand on the shoulders of giants', I think it is important nonetheless to do such studies.

Before proceeding to report the ERP experiments in Chapters 6-8, it is first necessary to describe the preparatory tests for these experiments; this is the topic of the next chapter.