The resolution of the clause that is relative?

Prosody and plausibility as cues to RC attachment in English: Evidence from structural priming and event-related potentials.

Daniela Zahn, MSc, M.A.

Submitted in fulfillment of the requirements for the Degree of PhD

Institute of Neuroscience and Psychology
College of Science and Engineering
University of Glasgow

20th of December 2013

Acknowledgements

I would like to take this opportunity to thank a couple of people.

Firstly, I am very grateful to Dr. Christoph Scheepers for his continued support, advice and help. Thank you to Dr. Guillaume Rousselet for the time he spent to make Matlab and EEG a less daunting experience.

Thanks goes to Dr. Sibylle Mohr, Dr. Dominic Thompson, Dr. Sara Bögels and Dr. Isabel Orenes for being great office mates (in varying combinations).

My greatest thank you, of course, goes to my boys. Saami for not minding my absence too much - Hail Minecraft!
Alberto, without your patience, love and the occasional kick in the b*** I could not have done it. I am forever in your debt.

This thesis is dedicated to my late parents, Ulrike Zahn und Dr. Gunther Zahn. Thank you for installing into me an eye for detail, curiosity and a sceptic mind.

Author's Declaration

I declare that this thesis is my own work carried out under normal terms of supervision.

Abstract

In spoken language, different types of linguistic information are used by the parser to arrive at a coherent syntactic interpretation of the input.

In this thesis I investigated two of these information sources, namely overt prosodic features and plausibility constraints. More specifically, I was interested in how these cues interact in resolving the relative clause attachment ambiguity.

Much research has explored the single cues and much is know about the influence that each cue exerts independent of the other. However, the interaction of prosodic and semantic cues to attachment has to date received little attention.

Two experimental paradigms were used in 4 experiments, i.e. the offline method of structural priming and the online method of event-related potentials.

The data from these experiments suggest that the cues interact in a complex way. The results imply that the prominence of the dispreferred cues, the surprisal and the type of revision associated with them play a major role during processing. I propose three processing principles that might account for the results observed.

Table of Contents

1.1. The challenge

One of the best known examples on the pitfalls of punctuation is the following taken from Truss (2003):

1. The panda eats shoots and leaves.

This sentences tells us something about a panda's diet. However, punctuation changes the meaning completely as in (2):

2. The panda eats, shoots and leaves.

The same string of words now describes what the panda has been doing during the day. The example epitomises how vague language can be and the problems that might arise from this. So, how can we make sense of the intended meaning?

The magic of language is, it is inherently ambiguous. Ambiguity is created because language is multidimensional. It is commonly assumed that to be able to comprehend what is being communicated a comprehender must integrate information from basically 2 sources: syntax and semantics. Syntax refers to the building block a language is made up of such as nouns, verbs etc. and how they are put together. Semantics refers to the meaning of words and how they are combined.

However, as the example above shows clearly, there is at least one other source of linguistic information that needs to be considered. As mentioned, a different punctuation pattern changes the meaning of the string of words above. Commas, full stops and question marks (etc.) give us an indication of how the string of words should be grouped together to understand its meaning.

For example, the ambiguity arises at the verb *eats*. On the one hand, *eats* could be interpreted as the transitive like (1), i.e. it needs to be followed by a direct object (*The panda eats what?*). On the other hand, the verb *eats* could be interpreted as the intransitive like (2) (i.e. *The panda eats and then does something else.*). Thus, the presence or absence of a comma guides the interpretation, i.e. the integration of syntactic and semantic information at this point.

Interestingly, punctuation is actually the graphic representation of what is called intonation or prosody. Prosody is a feature of spoken language. It describes the pattern of pauses and the changes in pitch over the course of a sentence or a number of sentences as in the example above. Thus, a listener must integrate these three sources of linguistic information to make sense of what is being said.

How this integration works, i.e. how humans process the prosodic, semantic and syntactic components of language, is a matter of much debate. Since ambiguities are abundant in language they present an excellent testing ground to study the contribution of different linguistic information sources such as plausibility and prosody to the resolution of said ambiguities in a controlled way.

1.2. Thesis overview

The question my PhD research was addressing, was: How do prosodic and plausibility information interact to resolve the temporary syntactic ambiguity present in relative clause attachment?

To answer this question, I investigated relative clause attachment from two points of view, namely from a psycholinguistic and a neurocognitive one.

Firstly, in **Chapter 2** I will give a general introduction to the field of ambiguity resolution. I briefly review different psycholinguistic approaches to ambiguity resolution before turning to challenges the approaches face. Based on this analysis, I will motivate the choice of ambiguity under consideration in this thesis, i.e. the relative clause (RC) attachment ambiguity. I will then turn to psycholinguistic models of RC ambiguity resolution. Subsequently, an introduction into neurocognitive models of language processing is given and I will summarise electrophysiological findings on language processing in general and RC attachment specifically.

In **Chapter 3** I will introduce and evaluate the research methodology used in the present experiments. The methods discussed are structural priming and event-related potentials (ERPs).

Chapter 4 comprises an exploratory pilot study that uses the priming paradigm for the first time. Prosodic and plausibility cues to RC attachment were used in a crossed design to investigate their interaction. The aim of the pilot study was to get a first indication of what the interaction of prosody and plausibility in RC attachment might look like and to generate more accurate predictions for further experiments.

Chapter 5 gives a detail overview of the materials used in the main experiments of this thesis. I will present the results of the acoustic, and auditory-perceptual validation of the prosodic manipulation, as well as the analysis and validation of the plausibility constraints employed in the items.

In **Chapter 6** the priming paradigm is used to investigate the interaction of prosody and plausibility in more detail. Additional subject-specific variables were recorded to evaluate possible individual differences that might play a role during the resolution of RC attachment. These variables included acoustic discrimination ability and listening memory span. The analysis of the main priming experiment 1 returned some unexpected results, namely two clusters of participants that displayed opposing attachment behaviour. Therefore, a second priming experiment was run that was a direct replication of the first main priming experiment. This is detailed **in Chapter 7**. Since both main priming experiments revealed very similar results it was decided to combine the data of these experiments. **Chapter 8** summarises the combined analysis of main priming Experiment 1 and 2.

In **Chapter 9** the interaction of prosody and plausibility was investigated using electrophysiological measure of event-related potentials.

Finally, **Chapter 10** concludes with a summary of the findings, final remarks, and future directions.

Please note that all sentence examples, figures and tables are numbered with each single chapter.

2. Introduction

2.1. Ambiguity resolution– A short overview

In psycholinguistics, much of the research being done relies on the resolution of syntactic ambiguities (Van Gompel & Pickering, 2006). A sentence is syntactically ambiguous when it is unclear what the grammatical structure of the whole sentence or part of that sentence is. The syntactic ambiguity can be local, i.e. affect the interpretation of a part of a sentence or global, i.e. affect the interpretation of the whole sentence.
Consider a sentence like 1.

1. The woman examined by the doctor turned out to be seriously ill.

At the point of *examined* the sentence is temporarily syntactically ambiguous. In one analysis the verb *examined* could be interpreted as past tense main verb, therefore, as part of the main clause, i.e. *the woman examined something*. Thus, the next syntactic component should be a noun in direct object position (e.g. *the woman examined the evidence*). Alternatively, the verb *examined* could be interpreted as past participle and as part of a (reduced) relative clause, i.e. *the woman that was examined…* . Hence, the upcoming syntactic component could be an adverb or a prepositional phrase (e.g. *the woman (that was) examined regularly / at the hospital*). It also implies that the main verb of the sentence is still to be expected.
Most readers will experience difficulty once they encounter the prepositional phrase *by the doctor* in (1). This phrase syntactically disambiguates whether the verb *examined* is to be analysed as a main clause verb or as part of the reduced relative clause. Readers initially tend to adopt the main clause interpretation and then have to re-analyse when they encounter the prepositional phrase *by the doctor* to arrive at the correct

interpretation (Ferreira & Clifton, 1986; Rayner, Carlson, & Frazier, 1983). The fact that readers initially tend to adopt the main clause reading suggests that there is a preference for processing an ambiguity. So the question needs to be answered what guides this preference.

There is widespread agreement that at any given word in a sentence such as (1) linguistic information is incorporated into the current interpretation of a sentence "word-by-word", i.e. incrementally. However, there is much debate about the manner in which the steps are taken. I.e. is syntactic information of the word currently being processed analysed first and other information such as plausibility and (possibly prosody) processed later? Or, is all linguistic information available immediately?

Readers tend to adopt one syntactic interpretation first, like the main clause interpretation in (1). This would argue for a *serial* manner of processing. The first step is taken by committing to a syntactic main clause analysis. No further linguistic information is taken into account at this point. The second step is taken when the initial analysis turns out to be wrong because the subsequent linguistic information is incompatible with the initial analysis. Thus, reanalysis ensues. Models that adopted this assumption would be serial, syntax first models. These are also called *two-stage accounts.*

The most influential two-stage account was proposed by Lyn Frazier (1979, 1987) and became known as the "Garden-Path" theory. In its simplest form, the model is based on the notion of building phrase-structures, e.g. S -> NP VP, VP -> V PP etc. Building a phrase-structure expresses the idea that a sentence (S) consists of a noun phrase (NP) and a verb phrase (VP). In turn, a verb phrase (VP) can consist of a verb (V) and a prepositional phrase (PP) etc. Consequently, these phrase structures can be combined into "sentence trees" by sticking them together. In case of an ambiguity, the processor has to decide which analysis to pursue based on certain heuristics. The heuristics would guide the parser so that only one syntactic interpretation is being followed at any given time. The presumed existence of processing heuristics is grounded in the assumption that the human parser has a limited ability to compute input. The limitations are imposed by the general cognitive architecture, most notably by working memory. That is, a single analysis is pursued because multiple analyses at the same time require more memory resources and are therefore very demanding.

One of the heuristics that Frazier proposed for ambiguity resolution is called *Minimal Attachment*. It assumes that an ambiguous phrase is attached to the preceding tree structure assuming the simplest possible structure. In the example above (1) minimal attachment would predict that readers initially interpret the verb *examined* as the main clause verb because it would involve a simpler structure than assuming that *The woman examined* is part of a more complex noun phrase. When encountering the prepositional phrase *by the doctor* it becomes clear that the main clause interpretation is incorrect. So the tree needs to be revised. The second heuristic Frazier introduced is called *Late Closure*. It states that incoming material should be attached into the constituent currently being processed. This means that incoming information should form part of the current phrase rather than start a new phrase. Take (2 a and b):

(2) *a. Since Jay always jogs a mile this seems like a short distance to him.*
b. Since Jay always jogs a mile seems like a very short distance to him.

In (2) the noun phrase *a mile* could either be interpreted as the direct object of the verb *jogs* (2 a) or as the subject of the main clause (2 b). The *Late closure* heuristic would suggest that people interpret the noun phrase as direct object of *jogs* (2a) as this is the current constituent. In contrast (2 b) would be dispreferred because a new constituent gets introduced. Indeed, Frazier & Clifton (1982) found longer reading times for sentences such as (2 b) compared to (2 a) which was taken as evidence that *Late closure* was exerting influence during processing.

Early experimental evidence has provided support for two-stage models (e.g., Frazier & Rayner, 1982; Frazier, 1987; Rayner et al., 1983; Ferreira & Clifton, 1986). Crucially, these accounts assume the primacy of syntax. Syntactic information is processed first and all other linguistic information like semantics or prosody will be considered at a later stage. This means that prosodic or plausibility information can reinforce or reject the initial syntactic analysis which is guided by purely syntactic heuristics such as *Minimal attachment* and *Late closure*. Thus, prosodic or plausibility information would trigger reanalysis by falsifying the initial syntactic structure. But prosodic or plausibility information does not influence the initial structure building itself. That is why these models are also called syntax-first models.

A number of studies have challenged syntax-first, two-stage models. McRae, Spivey-Knowlton and Tanenhaus (1998) firstly used a sentence completion task to look at the influence of thematic fit. Thematic fit in their interpretation refers to the likeliness of a noun to be an agent or a patient during an action. Using structures analogous to (1), fragments such as (3 a) were completed more often with a reduced relative clause than sentences such as (3 b)

(3) *a. The crook/arrested/ by/the detective*
b. The cop/arrested /by/the detective

They concluded that a typical patient (*crook*) of an action such *arresting* is more likely to be interpreted as a reduced RC than a typical agent (*cop)* of that same action. Also, a self-paced reading task supported this conclusion. Completed materials based on (3) showed shorter reading time for typical patients then for typical agents. In sum, their results indicated that information such as thematic fit of a noun is taken into account during processing right from the start. This also suggests that readers process not just syntactic but different types of information in *parallel* and without delay. Interactive models or often-called constraint-based (or constraint-satisfaction) models such as McRae, Spivey-Knowlton and Tanenhaus (1998) postulate that all possible interpretations of a current structure are activated in parallel. The syntactic structure that receives most support by the various constraints on a word-by-word basis ultimately wins the competition for the best overall analysis (Spivey & Tanenhaus, 1998; McRae et al., 1998; Spivey & Sedivy, 1995). Some of the constraints that could influence ambiguity resolution have been identified. They include different types of frequency information (McRae et al., 1998), plausibility related features such as animacy and concreteness (van Gompel et al., 2005; Desmet et al., 2006) as well as context information (Spivey & Tanenhaus, 1998). However, whether there is an effect of plausibility on initial syntactic processing or not is still debated as there is also evidence against it (Clifton et al., 2002; Ferreira & Clifton, 1986). The problem that interactive parallel models face is that it is very difficult to specify all the constraints that apply during ambiguity resolution. Without that knowledge it is difficult to falsify such a model as it could always be argued that another, unknown constraint exerts influence. It would just be a matter of finding that constraint. Also note that the above evidence does not necessarily prove syntax-first models wrong, because the latter could simply assume that constraints such as thematic fit would modulate reanalysis difficulty.

Constraint-satisfaction models receive support from studies into prosody and ambiguity resolution. Prosody refers to the rhythm, stress, and intonation of speech. Here, I will primarily focus on one particular aspect of overt prosody, which is roughly characterised by pauses (boundaries) and changes in fundamental frequency (F0) over the course of a spoken sentence. Indeed, this aspect of prosody is closely associated with, but not identical to syntax in that both interact in grouping blocks of meaning together in phrases. Two levels of prosodic grouping are commonly assumed: an intermediate phrase (ip) and an intonational phrase (IP). The intermediate phrase (ip) is the first prosodic unit above the prosodic word; each ip must include at least one pitch accent and ends with a boundary tone, either [H-] or [L-]. The intonational phrase (IPh) must contain at least one ip and ends with a stronger boundary tone; the IP boundary itself is denoted IPh (Pierrehumbert, 1980).[1] The tone assignments are related to

[1] Pierrehumbert commonly adopted notation denotes pitch accents as either L (for low) or H (for (for high) in the speaker's pitch repertoire. [L-] signifies a low pitch accent that occurs at the boundary of an ip; [L%] and [H%] denote strong boundary tones found at the end of an IP (IPh).

phonetic properties of the actual speech signal. There are parameters that indicate a boundary tone such as a drop in F0 and an increase in pre-boundary syllable duration (House, 1990; Klatt, & Cooper,1975; Klatt,, 1976). These features are inherent in speech, manifesting themselves in changes of fundamental frequency, intensity and duration over the course of a spoken sentence. As such, they mark the boundaries of linguistically meaningful units as well as their prominence.

I will primarily focus on one particular aspect of overt prosody namely prosodic boundaries or pauses. It has been shown that prosody, especially prosodic boundaries (pauses), can influence the resolution of temporary syntactic ambiguities (Lehiste, 1973; Price et al., 1991; Speer et al. 1996). Take (5) (from Speer et al. 1996):

(4) a. Whenever the guard checks [pause] the door is locked.
b. Whenever the guard checks the door [pause] it's locked.

This is a classic early vs. late closure ambiguity. It uses verbs that can be either transitive (take a direct object) or intransitive (not be followed by a direct object).

In (4) an ambiguity arises between the verb *checks* and the noun *door*. In (4 a) *door* is the direct object of *checks*. In (4 b) *checks* is the predicate of the subclause, thus *door* is subject of the main clause. However, at this point it is unclear which status *checks* has (transitive vs. intransitive) and consequently which role *door* fulfils (direct object vs. subject of main clause).

Kjelgaard and Speer (1999) showed that the insertion of a pause after *checks* reliably biases and facilitates the interpretation of *checks* as intransitive. Equally no pause after *checks* and a pause after *door* reliably biases towards the transitive interpretation. They also showed that a pause at the wrong place interferes with ambiguity resolution. Thus, prosodic information, precisely, the presence of a pause at points of syntactic uncertainty can influence processing very early on. Models that initially ignore any linguistic information that is not syntactic might find this difficult to explain.

To sum up, on the one hand, there is support for serial two-stage syntax first models. This support comes largely from studies looking at the syntactic means of disambiguating a temporary syntactic ambiguity. On the other hand, there is a substantial body of research that supports parallel constraint-satisfaction based processing of all available linguistic information without delay. The support is based on the resolution of syntactic ambiguities looking at syntactic, semantic or prosodic information.

2.2. General challenges in research of ambiguity resolution

Research into ambiguity resolution in general has provided us with a wealth of data. On the basis of this, detailed models of ambiguity resolution have been developed that have furthered our understanding of language processing. In spite of this progress, the psycholinguistic literature gives the impression of being fragmented and challenging to reconcile. These challenges stem from three sources. Firstly, there is a wide range of research questions that are being addressed. Naturally, this leads to the necessity to limit experimental parameters and hypotheses. Secondly, there is a broad range of experimental paradigms being used. Each paradigm has its merits, thus it is suited to produce results that need to be carefully interpreted with the limits of the experimental paradigm. And thirdly, as stated before, there are numerous syntactic ambiguities to choose from. Ambiguities share a common feature, i.e. at some point in the sentence it is unclear what the grammatical structure is. Nonetheless, ambiguities also differ substantially. I will briefly address each in turn.

Firstly, one community of research seems to concentrate on syntactic and semantic factors during processing. It is trying to address questions like (Pickering & van Gompel, 2006):

" […] what guides initial processing? Why does processing difficulty occur? How difficult is reanalysis? Do people construct ungrammatical representations?" (Pickering & van Gompel, 2006, page 4, 26, 30, 39)

Studies here concentrate on determining the time course of processing, i.e. when certain information is used and what the underlying representations for processing are. The linguistic information referred to is almost exclusively syntactic and / or semantic in nature (Ferreira & Clifton, 1986; Spivey & Sedivy, 1995; Spivey & Tanenhaus, 1998; McRae et al., 1998; Clifton et al., 2003; van Gompel et al., 2005; Desmet et al., 2006). Take McRae et al. (1998) reported here as 5 as an example.

(5) *a. The cop arrested by the detective was guilty of taking bribes.*
b. The crook arrested by the detective was guilty of taking bribes.

They manipulated a semantic constraint like thematic fit of the agent (*cop* v. *crook*). They were interested in whether this semantic feature influenced processing of the structural ambiguity (main clause v. reduced relative clause reading) present at the verb *arrest*. Thus, they seemed to concentrate on the syntax – semantics interface.

On the other hand, a second group of research seems to concentrates on the mapping of syntax and prosody. Researchers try to determine what the relationship between syntactic and intonational form might be and whether one could be predicted from the other. Usually this is done by looking at the production side (Speer et al., 2003, Carlson, 2009,) Consider (6 a v. b) as an example:

(6) *a. Susie learned that Bill telephoned after John visited.*
b. Susie learned that Bill telephoned [pause] after John visited.

Both sentences are globally structurally ambiguous. It is not clear whether Susie learnt something after John visited or whether Bill telephoned after John visited. Crucially, both interpretations are equally plausible. The *Late* *closure* principle of the garden path model would predict that *after John visited* is preferentially attached to the second VP *telephoned.* However, the insertion of the pause before the ambiguous prepositional phrase (6b *after John visited*) tends to bias listeners towards the first interpretation. Whereas the absence of a pause at this point seems to support the latter interpretation where *after John visited* modifies *learned*.

The goal of studies like these seemed to be to understand the mapping of prosody and syntax. Thus, they seem to concentrate on the syntax – prosody interface. Experimental research relies on constraining the questions one asks so predictions can be made and conclusions can be drawn. Therefore, factors of interest such as syntax and semantics or syntax and prosody, are isolated in carefully selected utterances. The challenge with language research, however, is that the basic building block like syntax, semantics and prosody do not exist in isolation, especially in spoken language[2]. Therefore, the subsequent step should be to look at syntax, semantics and prosody together.

The second challenge for research of ambiguity resolution concerns the research methodology. To study ambiguity resolution, different research paradigms are commonly used. Syntactic and semantic factors in ambiguity resolution have been studied extensively via recording of eye movements during reading (Frazier & Rayner, 1982; Rayner et al. 1983; Spivey & Tanenhaus 1998). Often a combination of methods is used, e.g. eye movements and self-paced reading (Ferreira & Clifton, 1986), sentence fragment completions and self-paced reading (McRae et al. 1998) or analyses of text corpora, sentence fragment completions and self-paced reading (Spivey & Sedivy, 1995; Scheepers, 2003).

Eye tracking as well as self-paced reading are very useful in determining the ease or difficulty of readers' comprehension of a text (Staub & Rayner, 2007). Both are considered an on-line measure, hence they are suited to look at the time course of processing, i.e. when certain information is used.

[2] This is most likely also true for written language. See Frazier, 2002 on reading and implicit prosody.

Sentence fragment completions and the analyses of text corpora are suited to study the underlying representations for syntactic processing.

Prosodic factors in ambiguity resolution have been studied using some kind of judgment of either the naturalness of the prosodic structure of an utterance (Frazier, Clifton & Carlson, 2004) or grammaticality (Kjelgaard & Speer, 1999). Very common is also what is called forced choice comprehension and cross-modal naming (Kjelgaard & Speer, 1999; Carlson, Frazier, Clifton 2009; Carlson, Clifton, & Frazier, 2009; Snedeker & Casserly, 2011). Both of these methods are based on the idea that participants make a choice based on the linguistic information presented in the stimulus. Thus, they produce data that gives an indication of how participants interpret the stimulus and thereby enable conclusions about the function of prosody in processing but not necessarily when a prosodic cue comes into play. Another on-line method like eye-tracking is rarely used (Weber, Grice & Crocker, 2006).

The methods used in research are diverse and suited to answer the question at hand. However, the diversity makes it problematic to link and compare the results they produce. Consequently, it seems beneficial to employ a combination of off- and on-line methodologies already used to study either syntactic and semantic or prosodic ambiguity resolution. The results can then be combined to inform models of ambiguity resolution.

The third challenge is, that there are numerous syntactic ambiguities to choose from. Much research on syntactic ambiguity resolution has used the reduced relative clause vs. main clause ambiguity like (1) (Frazier & Rayner, 1982; Rayner et al. 1983; Spivey & Tanenhaus 1998, McRae, 1998).

In contrast, in prosody research the VP1 v. VP2 attachment ambiguity (as in (6)) has been extensively used (Ferreira & Clifton, 1986; Snedeker & Trueswell, 2003; Carlson, Clifton, & Lyn Frazier, 2009; Breen, Watson & Gibson, 2010). Also early v. late closure ambiguities like (2) (Speer et al. 1996; Kjelgaard & Speer, 1999) are used.

Although these ambiguities share the common feature that at some point in the sentence it is unclear what the grammatical structure is, they also differ.

The difference has mainly to do with the part of the sentence where the ambiguity occurs. I.e. take (1) and (4) repeated here as (7) and (8):

(7) The woman examined by the doctor turned out to be seriously ill.

(8) a. Whenever the guard checks [pause] the door is locked.
b. Whenever the guard checks the door [pause] it's locked.

In the case of the reduced relative clause vs. main clause ambiguity (7) and the early v. late closure ambiguity (8), so called core arguments of a sentence are involved. Core arguments of a sentence are the basic constituents that are needed to make a sentence grammatical, for example a subject, a verb (predicate) and possibly an object.
In both cases above the ambiguity concerns the verb (the predicate) and its relationship with either the subject (7) or the direct object (8). So to make sense of the core grammatical structure, it is absolutely essential to solve the ambiguity.
In contrast, consider (6) here repeated as (9):

(9) *Susie learned that Bill telephoned after John left.* (Carlson et al., 2001)

The ambiguity here is a so-called attachment ambiguity. An attachment ambiguity involves a constituent of the sentence that is not part of the core argument structure but rather modifies it[3]. Thus, it should be no problem to make sense of the core grammatical structure of the main clause in (9) (*Susie learnt that Bill telephoned)*. However, it is temporarily unclear which core constituent is being modified in (9). The prepositional phrase in (9) modifies either *Susie learnt something after John left* or *Bill telephoned after John left*.
It has been argued that there is a qualitative difference between ambiguities that involve core arguments and those that do not (Frazier & Clifton, 1997) which might impact processing of said ambiguities. Indeed, Frazier & Clifton

[3] I will discuss attachment ambiguities in more detail under 2.4.1.

(1997)'s *Construal Theory* postulates that different principles apply to the processing of modifiers versus core arguments which could be seen as "softening" the universality of the garden path model. On these grounds it might be difficult to compare conclusions drawn from experiments that use different ambiguities.

To sum up, there seemed to be three main challenges in research of ambiguity resolution. Firstly, there is a seeming divide in research between the processing of syntax and semantics and the processing syntax and prosody. Because syntax, semantics and prosody don't exist in isolation in spoken language, it might be problematic to develop models that include all three if the interaction of all three is not addressed explicitly. Secondly, different strands of research seem to favour different methods, which makes the results difficult to compare. And thirdly, an array of ambiguities is being used that differ in type and possibly in quality of processing.

How can these challenges be addressed? Unfortunately is beyond the scope of this thesis to tackle all three challenges at the same time. However, a first step could be to choose an ambiguity in which the influence of syntax, semantics and prosody in isolation has been studied before, e.g. the relative clause attachment ambiguity. Thus all three can be combined to look at the interaction. Also, off-line and on-line methods of research can be employed to address the question of time course and constraints on processing.

2.3. The syntactic ambiguity under consideration – relative clause attachment

Attachment ambiguities are a class of syntactic ambiguities that involve a constituent of the sentence that is not part of the core argument structure but rather modifies it. Attachment ambiguities are suited to study language processing because when an ambiguity occurs it needs to be resolved in order to make sense of the linguistic input. Thus, how the ambiguity is resolved might give us an idea of kinds of constraints that underlie language processing. It might also shed light on when different sources of linguistic information are combined during processing. The relative clause attachment ambiguity is very well suited to study syntactic ambiguity resolution because various factors that influence the resolution have been investigated before.

2.3.1. The relative clause attachment ambiguity – a brief overview

One frequently studied example of an attachment ambiguity is the relative clause attachment ambiguity. It is found in sentences comprising a complex noun phrase (NP of NP) with an adjacent relative clause (RC), as in (10).

(10). *Someone shot the servant of the actress who was almost deaf.*

The main clause contains a complex noun phrase, i.e. *the servant of the actress*. This complex noun phrase is modified by the following relative clause (RC) *who was almost deaf*. (10) is a global ambiguity because at this point it is unclear which of the two nouns is being modified. If the first noun is being modified (*the servant was almost deaf*) the relative clause attaches higher up in the syntactic tree (high-attachment, Figure 1). In contrast, if the second noun is being modified (*the actress was almost deaf*) the RC attaches lower down in the syntactic tree (low-attachment, Figure 2). Thus, a syntactic ambiguity is created.

Figure 1 Phrase-structure representation for high-attachment (HA) of a relative clause (RC)

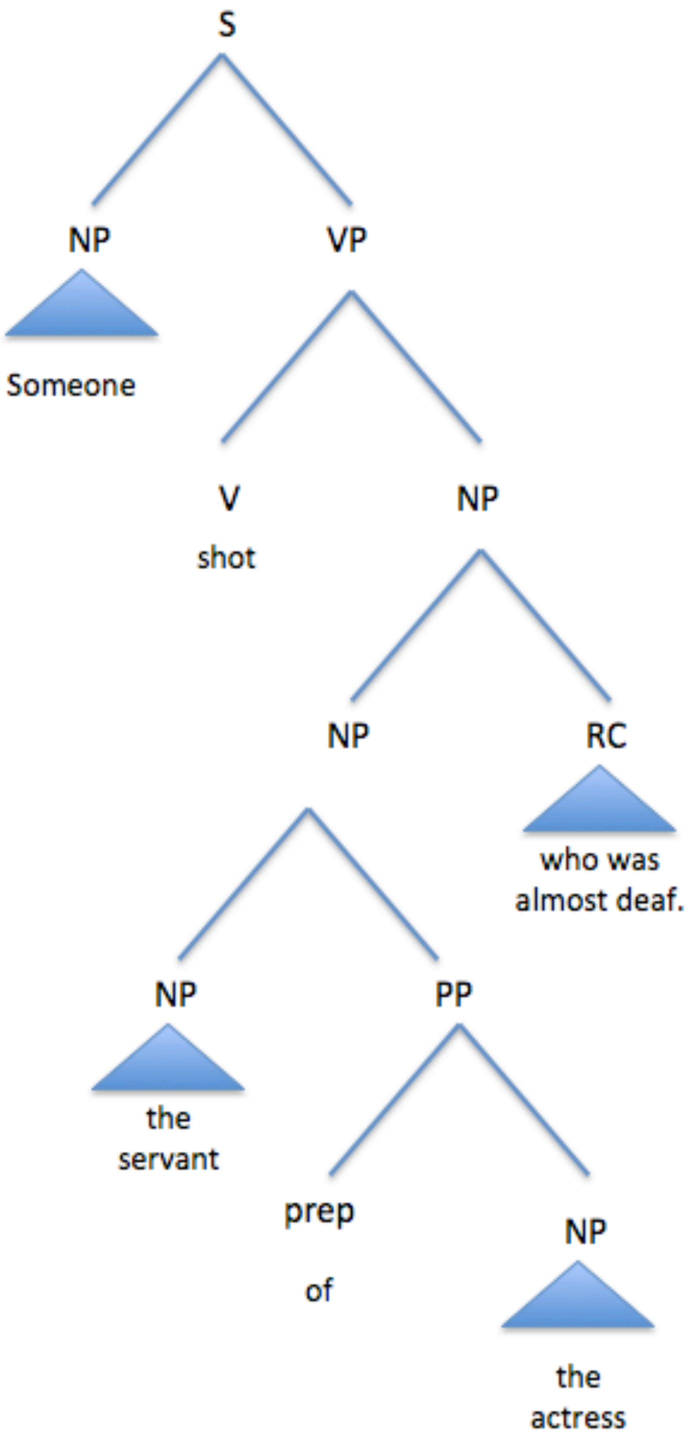
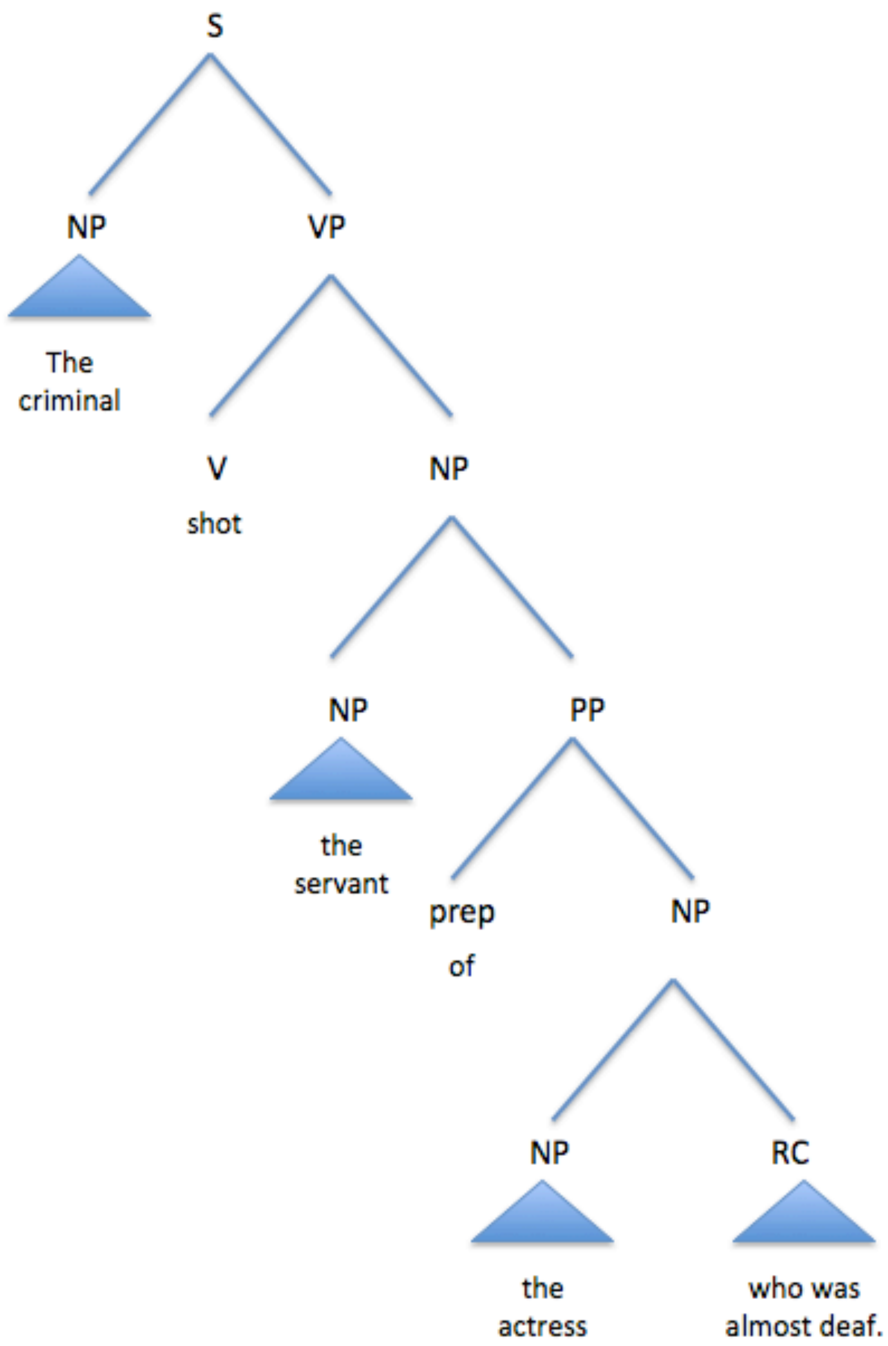


Figure 2 Phrase-structure representation for low-attachment (LA) of a relative clause (RC).



There are several ways of disambiguating a relative clause towards either one or the other interpretation (high attachment vs. low attachment). Firstly, syntactic means can be employed like (11 a and b):

(11) a. *Someone shot the servants of the actress who were almost deaf.*
b. *Someone shot the servants of the actress who was almost deaf.*

The number agreement between the attachment host (*servants* v. *actress*) and the verb in the relative clause disambiguates attachment towards high attachment in (11 a) and low attachment in (11 b). This type of disambiguation has been used by Kaan and Swaab (2003) for example.
Secondly, general plausibility constraints can be applied. For example (12 a and b):

(12) a. *Someone shot the servant of the actress who was very servile.*
b. *Someone shot the servant of the actress who was very famous.*

The final lexical-semantic information implies whether the RC modifies NP1 (*the servant*) or NP2 (*the actress*). Based on plausibility constraints in (12 a), i.e. *the servant* is more likely to be *servile* than the actress, the relative clause attaches higher up in the syntactic tree. By contrast, in (12 b) the relative clause is said to modify the more recent noun NP2, i.e. *the actress* and attaches lower down in the syntactic tree. This is based on the notion that an actress is more likely to be famous than a servant.
Different aspects of plausibility information (e.g. animacy of the host noun phrases) can constrain the processing of the sentence in such a way that one interpretation (implying either high- or low-attachment of the relative clause) is more salient than the other. Indeed, in research on reading, plausibility constraints similar to (12 a and b) have frequently been used to maximally disambiguate relative-clause attachments in complex noun phrase structures (e.g. Cuetos & Mitchell, 1988; Carreiras & Clifton 1993; Traxler, Pickering &

Clifton, 1998; van Gompel, Pickering & Traxler, 2001; van Gompel et al., 2005). It seems that – in English at least – such constraints provide a very effective cue to the final high- or low-attachment interpretation of the relative clause. Thirdly, an RC attachment ambiguity can be resolved by means of prosody. Most relevant in the present context is a study by Clifton., Carlson, and Frazier (2002) who investigated, among other things, the influence of overt prosodic boundaries (pauses) on the comprehension of relative-clause attachment ambiguities such as (13 a and b).

(13) a. *Someone shot the servant of the actress [pause] who was almost deaf.*
b. *Someone shot the servant of the actress who was almost deaf.*

Again, the sentence is globally ambiguous and is not disambiguated by means of syntax or semantics. Thus, it is impossible to use plausibility constraints to determine attachment of the RC. Remember, a syntactic principle like *Late closure* would predict a low attachment preference. However, prosodic means might exert influence on processing in this case. That is, the insertion of a pause before the relative clause (13 a) would be expected to trigger high attachment of the RC. This is based on the idea that the pause blocks attachment to the most recent noun NP2, thereby forcing the attachment of the RC to the next available noun NP1. This idea would imply that a syntactic principle like *Late closure* could be overridden by prosodic means. In the garden path model, this would take place during the revision stage. The absence of the pause (13 b), on the other hand, is expected to trigger low attachment to NP2 because it is the most recent noun and attachment to the most recent noun should be easiest to achieve. This would imply that the absence of a pause is somehow linked to a syntactic principles like *Late closure* and *Minimal attachment* and thereby confirm the initial low attachment analysis.

Using an offline comprehension task, Clifton et al. (2002) found that an “informative” pause (i.e. a prosodic boundary larger in size or prominence than any preceding prosodic boundary) before the relative clause in a spoken sentence such as (9 a) reliably biased listeners to assume a high-attachment interpretation of the relative clause (Clifton et al., 2002, Experiment 2).

2.3.2. Relative clause attachment and models of ambiguity resolution

There are different ways of disambiguating a global RC attachment ambiguity such as (10). Syntactic (11), plausibility (12) or prosodic (13) means can be employed. However, it has become clear that listeners or readers are usually not aware of the ambiguity in (10) and they display a preference for the resolution of the ambiguity. Native speakers of English tend to display a low attachment preference for globally ambiguous sentences. This was first reported by Cuetos & Mitchell (1988) who found a low attachment preference for English.[4] In the first instance this would argue for universal processing heuristics such as *Late Closure* and *Minimal Attachment* proposed by Frazier under her "Garden Path" theory (1979, 1987a). The *Late Closure* principle states that incoming material should be attached into the constituent currently being processed. This means that incoming information should form part of the current phrase rather than start a new phrase. Crucially, a heuristic like *Late Closure* was assumed to be universal and language independent. Cuetos & Mitchell's (1988) findings called this assumption into question because they also reported a high attachment preference for Spanish. If *Late Closure* and *Minimal Attachment* were truly universal, that apply to all languages, Spanish would display the same low attachment preference as English. However, this is not the case. Consequently, much research has been done into cross-linguistic attachment preferences, e.g. Spanish (Carreiras & Clifton 1993; Gilboy, Sopena, Clifton & Frazier 1995), Italian (De Vincenzi & Job, 1995), Dutch (Desmet, Brysbaert & De Baecke, 2002; Desmet et al., 2006), and German (Hemforth, Konieczny, & Scheepers, 2000; Scheepers, 2003).

[4] More evidence, however suggests that English shows either a weak preference for low attachment or no clear preference at all (e.g., Carreiras & Clifton, 1993, 1999; Traxler, Pickering, & Clifton, 1998).

One of the aims of this type of research was to determine to what degree processing heuristics such as *Late Closure* and *Minimal Attachment* are universal or whether different models of language processing might offer a more suitable account for the data recorded.

Unfortunately, psycholinguistic research on RC attachment in the prosodic domain is limited. Some research has focused on the production side, i.e. the default phrasing (Jun, 2003). To date I am aware of only one behavioural study on the influence of pauses on the comprehension of RC attachment in English (Clifton, 2002). They found that the presence of a pause reliably biased listeners towards high-attachment of the relative clause. However, their research was concerned with the question whether listeners interpret prosodic boundaries as a local or a global event. They concluded that listeners take into account all preceding prosodic boundaries and interpret them relative to one another. The finding that the pause biased attachment could be explained by both models of ambiguity resolution. On the one hand, serial, syntax first approaches could argue that the influence of the pause is exerted via revision. On the other hand, parallel, constraint-satisfaction model could argue that the prosodic cue is taken into account immediately. It would make a high attachment interpretation more likely by adding support for this structure and no revision would be necessary.

2.3.3. Challenges in research of the RC attachment ambiguity resolution

While looking at ambiguity resolution in general I've noted a number of challenges.

The first challenge tackles the somewhat fragmented literature on RC attachment resolution. There is a wide range of research questions being asked. This in turn, constrains the parameters of the research undertaken and the theoretical modelling being done. When considering the RC attachment ambiguity it becomes apparent that the questions addressed concentrate on the plausibility – syntax interaction (Van Gompel, Pickering and Traxler, 2000) during processing. Despite their differences, there is a commonality in all models on RC attachment, be they serial, parallel or a combination of the two. That is, the resolution of the RC ambiguity has often been discussed in terms of the semantics / plausibility and syntax interface. RC attachment has also been used to test predictions of different types of models directly (van Gompel et al., 2005) which has furthered our understanding of strengths and weaknesses of different models. A (possibly unintentional) side effect of this is that the influence of prosodic factors has not been investigated in this context.

Then again, research on prosody in RC attachment resolution focuses on the prosody – syntax mapping (Clifton, 2002). Although recent results could lend support to parallel processing it seems not to be discussed in these terms. Rather, it is discussed in terms of the prosody – syntax mapping, i.e. what information a boundary or pause implies about the syntactic structure of the sentence (Clifton, 2002).

The second challenge arises from the diversity of research paradigms being used. The most frequently used paradigm to look at plausibility constraints on RC attachment is eye-tracking (Carreiras & Clifton, 1993, 1999; van Gompel, Pickering and Traxler, 2000; van Gompel, 2005). Sometimes eye-tracking is used in combination with questionnaire studies and explicit comprehension questions (Gilboy et al., 1995). As detailed above (2.3.) results produced from eye-tracking experiments provide an excellent source to study on-line comprehension. Thus, they are well suited to study the RC attachment ambiguity in the reading domain.

In research on prosody and RC attachment explicit comprehension questions were asked (Clifton 2002). The effects of prosodic manipulations tend to be very small and susceptible to task effects (Lee & Watson, 2011). Especially explicit comprehension questions seem to bias participants responses. More importantly, research in the auditory domain (Clifton et al. 2002) mostly investigated the influence of overt prosodic constraints on syntactic attachment under 'neutral' plausibility conditions (equal semantic support for either attachment alternative) and in the reading literature (e.g. van Gompel et al., 2005), plausibility constraints were often used to maximally disambiguate relative-clause attachments while keeping (implicit) prosodic constraints more or less constant.

To conclude, the divide assumed in the general literature on ambiguity resolution appears also in the literature on the RC attachment ambiguity resolution. On the one hand, processing of the resolution of RC attachment is being modelled with a special focus on the types of plausibility constraints involved. Again, by concentrating on this, the influence of prosodic factors is being somewhat overlooked. On the other hand, the influence of prosodic factors is studied to determine the mapping of prosody and syntax. Again, by concentrating on this, the influence of plausibility factors is being somewhat overlooked. The research methodology (eye-tracking vs. explicit comprehension) and materials (neutral prosody vs. neutral plausibility) make it difficult to combine the respective results and to evaluate their relative contribution. However, since plausibility and prosodic constraints on RC attachment have been studied in isolation before it should be possible to combine them. The chance to combine both factors in this type of ambiguity makes it ideal to look at the interaction of prosody and plausibility during processing.

Additionally, it is necessary to employ research paradigms that are unbiased, especially with regards to the influence of prosodic manipulations. But also research paradigms are needed that offer a window into on-line processing of the resolution of the RC attachment ambiguity. Hence, priming and event-related potential were chosen (a more detailed introduction to the methodology will be given in Chapter 3).

2.4. Ambiguity resolution – Neurophysiological evidence

Over the past 30 years research into online processing of language has done a big leap forward by employing neurophysiological measures of brain activity such as electroencephalography (EEG). During EEG recording, recurring electrical brain activity is associated with repeatedly presented linguistic input. They are called Event-related Potentials (ERPs). A more thorough introduction into the basics of ERPs as a research method will be given in Chapter (3). Much research has identified and replicated a number of language-related ERP components, which I will now turn to.

2.4.1. Event related potentials and language processing

Different types of linguistic information such as syntax, semantics and prosody have been related to different ERP components and effects. I will discuss the most important in turn.

There is one component that has been linked to semantic processing in general. It is a negative going waveform that starts around 300ms after onset of a semantically anomalous word, reaches its peak around 400ms and returns to baseline around 500ms. This component was termed N400. It is usually found on centro-posterior electrode sites but its topography can vary with stimulus features and input modality (Kutas & Federmeier, 2011). It was first reported for words that do not fit the semantic context they were presented in, such as *I take my coffee with cream and dogs* (Kutas & Hillyard, 1980). Since then it was found for single word context (Holcomb, 1993), for sentence context (Kutas & Federmeier, 2000) and for discourse in general (van Berkum, Hagoort, & Brown, 1999). Thus, Federmeier and Kutas (1999) argued that, the N400 is not only sensitive to the semantics of a context, but seems to reflect the ease of accessing or pre-activation of information in semantic memory. That is, expectations about how a sentence is to be completed either semantically, morphologically, phonologically or orthographically, can be pre-activated by specific contextual-semantic cues (for a summary see Baggio & Hagoort, 2011).

There are (at least) two components[5] that have been linked to processing of syntax. One is a negativity that can also be found between 300-500ms after word onset with a more (left) anterior topography, hence termed Left Anterior Negativity. LAN effects have mainly been observed for morphosyntactic violations. For example *The ice cream was eat*. This construction requires a past participle (*eaten*). Instead the infinitive (*eat*) is presented, leading to a violation of the expected syntactic form. (Friederici et al., 1996).

The second supposedly syntax-related component is called P600. The P600 is a positivity that starts around 500ms after word onset and it lasts usually around

[5] For the time being, I will ignore the ELAN as some doubt has been cast on its functional interpretation based on issues concerning its reliability and validity (Steinhauer & Drury, 2012).

500ms. The P600 often follows a LAN and is found for ungrammatical continuations of the broadest sense. Hagoort (2003) lists among others, some of the violations used for different languages, e.g. phrase-structure violations, subcategorization violations, violations in the agreement of number, gender, and case (Hagoort, 2003). When looking at the sentences used, like *The ice cream was eat.,* it is apparent that the violations are truly ungrammatical. However, a P600 has been found not just for ungrammatical violation but also for well-formed sentences that vary in complexity (Kaan et al., 2000) or include violations akin to semantic verb–argument violations (animacy) (Kuperberg, 2007). More interestingly, a P600 was reported for temporarily ambiguous sentences (Osterhout et al., 1994; Van Berkum et al., 1999). The difference between outright violations and complexity / ambiguity seems to be reflected in topographical differences between the respective P600 effects observed. A P600 for outright violations was commonly observed at posterior electrode sites whereas a P600 for complex or ambiguous sentences was found at (right-) anterior electrode sites.

Relatively recently a new ERP component was discovered that is linked to processing a prosodic boundary (or pause). It was first reported by Steinhauer, Alter and Friederici (1999) who showed that clear positive wave was elicited by a prosodic boundary in the speech signal. They termed it closure positive shift (CPS). The CPS is found bilaterally with a centroparietal distribution (Bögels et al., 2011). There is some discussion as to the exact onset of this component. Steinhauer et al. (1999) chose the sentence onset whereas later studies chose the offset of the last word before a pause (Pauker et al., 2011) or the onset of the stressed syllable of the last word before the pause (Bögels, 2011; Bögels et al., 2013). Bögels (2011) motivates the syllable onset by pointing out that the last syllable carries the acoustic information associated with a boundary, i.e. pitch information of the boundary tone and pre-final lengthening. Nonetheless, to date it remains unclear what acoustic information is absolutely necessary to elicit a CPS. Agreement exists that the offset of the CPS is triggered by the onset of the first word after the pause (Bögels et al., 2011). It has also become clear, that the CPS is not an instantiation of another positivity such as the P600. Pannekamp, Toepel, Alter, Hahne and Friederici (2005) showed that no semantic, syntactic or segmental-phonetic information is needed to elicit this component. Additionally, the CPS has been found in different languages such as German (Steinhauer et al., 1999), Dutch (Bögels, 2011), English (Pauker et al., 2011) and Japanese (Müller et al. 2005). On these grounds it has been argued that the CPS truly reflects some kind of prosodic processing. Most likely it could be interpreted as a reflection of structuring and / or phrasing of the input (Bögels et al., 2011).

I will now turn to what ERPs have so far been reported for processing of the relative clause attachment ambiguity.

2.4.2. ERPs and the RC attachment ambiguity

Some previous ERP research has looked at relative clause attachment. In a reading study, Kaan and Swaab (2003) investigated RC attachment in English. They looked at complex noun phrases (NP1 prep NP2) with an attached RC .

They found a posterior positivity for non-preferred high attachment and for ungrammatical attachment compared to the preferred low attachment. There seemed to be no difference between ungrammatical RC attachment and dispreferred high attachment when the disambiguation was syntactic in nature (number agreement between verb and attachment host). Both elicit a posterior P600 compared to preferred low attachment. In addition, they used structurally simpler single NP constructions and compared them to more complex noun phrases (NP1 prep NP2). When comparing these, they found that the more complex and ambiguous 2 NP conditions elicited a more right anterior positivity compared to the simple NP conditions. Therefore, syntactic complexity and temporal ambiguity seemed to be taken into account during processing. This was reflected in an anterior positivity as a response to more complex ambiguous sentences compared to simpler unambiguous ones.

Furthermore, Carreiras et al. (2004) studied RC attachment in Spanish which has shown to be a high attachment preference language (Cuetos & Mitchell, 1988). In their reading study, disambiguation of RC attachment was achieved syntactically, i.e. number and gender agreement between the last word and the intended attachment host. At the point of disambiguation two consecutive positivities were found. One was a broadly and slightly frontal distributed positivity between 500-700ms that can be compared to Kaan & Swaab's (2003) anterior positivity which could be a response to complexity. The second was a more posterior positivity between 700-1000ms that seemed to resemble the posterior, revision and repair related P600 of Kaan & Swaab's (2003).

In a number of ERP experiments that focused on reading and listening, Augurzky (2005) investigated RC attachment in German. Most relevent for the current body of work is the listening experiment. During the listening experiment, prosodic cues (*NP1 [pause] NP2 RC* for low attachment v. *NP1 NP2 [pause] RC* for high attachment) and syntactic disambiguation (gender marking at the relative pronoun) were manipulated. One prosody related CPS was found at the position of the pause for prosodic high attachment cue (*NP1 NP2 [pause] RC*). In contrast, two CPS' were reported for the low attachment prosodic cue (*NP1 [pause] NP2 RC*), i.e. one at NP1 and one at NP2. However, the auditory materials might have included an acoustic confound in the low attachment prosody conditions, i.e. the pitch trajectory at NP2 was highly comparable between high and low attachment conditions. Therefore, the results were very difficult to interpret[6].

To sum up, little is known about the electrophysiological correlates of RC attachment processing in the auditory domain. Evidence from reading studies in English (Kaan and Swaab, 2003), Spanish (Carreiras et al., 2004) and German (Augurzky, 2005) suggest that firstly, the complexity of the noun phrase was taken into account as evidenced by an anterior positivity in the P600 time window. Secondly, a language-typical preference seemed to play a role (at least in English and Spanish) which was reflected in a late posterior positivity for dispreferred attachment (high in English and low in Spanish). A very similar late posterior positivity was found for ungrammatical disambiguation (Kaan and Swaab, 2003). Prosodic cues seemed to have been taken into account during processing of RC attachment in German (as evidenced e.g. by the CPS in Augurzky, 2005). I will detail predictions for the interaction of prosody and plausibility based on these previous studies in more detail in Chapter 9.

[6] For more details see Chapter 9.1.

2.4.3. Neurocognitive models of language processing

The discovery of components like LAN, N400, P600, CPS has helped to develop detailed neurocognitive models of language processing, two of which I will present in more detail. I have mentioned before (Chapter 2.1.) that models of ambiguity resolution and language processing in general can be grouped into two main categories. One category comprises serial, syntax-first models and the other parallel, constraint-satisfaction models. This distinction is also reflected in neurocognitive models. It is important to note at this point that the two models that I will summarise below try to account for language processing in general.
The most influential serial, syntax-first model was proposed by Friederici (2002). In her Neurocognitive model of auditory sentence processing she assumes that when the parser is presented with incoming information it initially constructs the simplest syntactic structure alone. The word category information of the word currently under consideration is retrieved from memory and held in working memory during processing. The structure building is supposed to be independent of lexical-semantic information. At a second stage, lexical-semantic and morphosyntactic information is retrieved from memory. The parser uses that information to assign a thematic role (who does what to whom etc.) to the input currently being held in working memory. Subsequently, a mapping of the role assignment onto the basic syntactic structure takes place. If this mapping is unsuccessful reanalysis and possibly repair need to be done for which general memory resources are recruited. Friederici also postulates a difference between outright violations of a structure and structural ambiguities. Ambiguous structures are processed slightly differently in that the initial structure building is not totally independent of non-structural information. Frequency of a particular structure or the semantic plausibility associated with the main verb are taken into account. Her model is syntax-first in the sense that syntactic phrase structure information is processed first, autonomously and prior to semantic information. Interaction of syntactic and semantic information is only observed during later stages of processing. Since specific types of information become available at specific times during processing, violations of a specific type of information should be detectable at certain times as well.

Consequently, Friederici (2002) aligns specific ERP components and their time windows with the above outlined processing stages. Language processing in the model consists of four distinct phases. During the first phase word category information is identified. Any violations at this point are reflected by the ELAN component (100-200ms). During the second stage (300-500ms), morphosyntactic and lexical-semantic information are assessed and integrated. Violations of morphosyntactic nature are reflected in the LAN component and violations of lexical-semantic constraints elicit a N400. If the integration is unsuccessful reanalysis and repair processes ensue which is reflected in the P600. At the time when the model was first published no specific prediction was made as to the function of prosodic factors evidenced by the CPS. In a more recent version of the model (Friederici, 2011), Friederici cautiously suggested that:

> "(…) [t]he prosody-syntax interaction may take place during different processing phases: *1*) during the initial phase of phrase structure building since the end of a syntactic phrase is marked prosodically, and/or *2*) during the second processing phase during which the verb argument structure is processed, since the constituent structure is also prosodically marked" (Friederici, 2011, p. 1383).

She supports this possible interaction with some of the above-cited ERP evidence (Bögels et al., 2010; Friederici, von Cramon & Kotz, 2007). Note, however, that no more clear-cut predictions as to the time course of the prosody-syntax interaction have been made in the model.

The second model I want to discuss is the Memory, Unification and Control (MUC) model proposed by Hagoort (2003, 2005). The MUC model assumes parallel, constraint-satisfaction based processing and it postulates three components, a memory repository, a unification buffer and a control apparatus (Baggio & Hagoort, 2011). The memory repository is something akin to a lexicon, i.e. it stores phonological, syntactic and semantic information associated with morphemes, words, and other constructions. Contrary to Friederici's model which seems to suggest that there is a difference between syntactic rules and lexical information stored in memory, Hagoort's memory repository stores what he calls unification ready structures. These structures are lexicalist in nature in as much as each word is associated with a structural frame and constraints on how frames can be combined (for more detail see Hagoort, 2005). Thus, chunks of syntactic structure are encoded on the word level and no syntactic rules need to be stored separately. Processing is consequently understood as a single process, namely unification. Much like any other model of language processing, word forms enter the unification buffer or workspace incrementally. The unification operation then consists of linking up lexicalist frames and checking for agreement of relevant features. This procedure is based on the constraints associated with the word currently being processed. Hagoort's model can be classified as parallel, constraint-satisfaction because it states that at any point during unification a number of alternative candidates for linking are available. Each one comes with its own set of constraints. During unification, attempts are made to solve on-line those sets of constraints by checking for agreement until the best option is chosen. Put simpler, some candidates have stronger unification links than others. This can for example be due to plausibility constraints that are associated with the syntactic frame. If one candidate carries more weight because of its plausibility constraints, it will bind with the preceding structure. The candidate also retains a context for subsequent stages of memory retrieval because of the constraints on binding it carries. Thus, unification is achieved. Finally, the control apparatus monitors executive functions such as turn taking (Baggio & Hagoort, 2011).

ERP components in Hagoort's model are interpreted differently. Firstly, the time window in which ERP components appear are not necessarily linked to accessing a specific type of information in a specific order as Friederici suggests. In Hagoort's model there are two main sources of unification difficulty. One is the failure to find a suitable candidate for a specific binding option. The other is an unsuccessful agreement check between the information a candidate carries and the constraints required by the binding option. An anterior negativity between 300-500ms like the LAN is supposed to be a reflection of both of these sources of difficulty. A more posterior negativity like the N400 is also associated with unification difficulty. The difficulty is due to combinatorial semantic processes and / or pre-activation of required information. Thus far however, Hagoort concedes that:

> "(...) it is still unclear what counts as a combinatorial operation from a brain systems perspective." (Baggio & Hagoort, 2011, p. 1353).

He suggests that the semantic combinatorial operations might be found at least at two levels, i.e. conceptual content and information structure. Information structure refers to the relative salience of information, e.g. marking something as new or given, topic or focus. Li, Hagoort, and Yang (2008) found that, in Chinese, the N400 to new accented information was larger than the N400 to new de-accented information. Hence, one way to explain an N400 would be that it reflects the necessity to recruit additional unification resources for information that is marked as more salient (Baggio & Hagoort, 2011). Pre-activation could be thought of as the flipside of this because "(...)in realistic processing situations, access, retrieval, pre-activation and unification are all part of word processing (...) (Baggio & Hagoort, 2011, p. 1361).

Another important ERP component that needs to be explained is the P600. As detailed above, a P600 has been reported for outright violations, complexity and ambiguity. Under the MUC framework, they seem to be treated very similarly. Violations, complexity and ambiguity are all associated with a certain degree of processing difficulty due to either ongoing competition between alternative unification options (ambiguity / complexity) or the search for a new unification option (violation). Consequently, the P600 reflects the time taken to establish unification links of sufficient strength. Amplitude and duration of the P600 are modulated by the amount of competition and, thus, the amount of processing difficulty. A P600 should always be elicited when unification is attempted (Hagoort, 2005).

Much like Friederici, Hagoort leaves the role of prosodic factors during unification somewhat underspecified. To date he has not explicitly incorporated the prosody related CPS. However, he links prosodic characteristics like boundary tones, pausing and lengthening to the idea of salience of information, specifically which aspects of an utterance get focused (Hagoort, 2003). At the same time he appears to imply that what gets focus in an utterance

> "(...) cannot be determined on the basis of information retrieved from memory, but requires an analysis of how lexical elements are unified into phonological structures spanning a multi word utterance. [Thus], in

> language comprehension syntactic, semantic and phonological unification processes operate concurrently and interact to some extent." (Hagoort, 2003, p. 418)

In sum, the distinction between serial, syntax-first and parallel, constraint-satisfaction models can be found in the literature on neurocognitive aspects of language processing. I have introduced 2 eminent models, namely, Friederici's Neurocognitive model of auditory sentence processing and Hagoort's Memory, Unification and Control model. Friederici assigns syntactic and semantic processing steps to ERP components and their time windows. Hagoort also relates specific ERP components to his model but gives them a very different functional interpretation.

Interestingly, both models leave a possible influence of prosody on semantic and syntactic processing rather vague. This leads to the question: what ERP evidence is there that prosody influences especially syntactic processing?

Over the past 15 years, research has been done to a great extent to address the role a prosodic boundary plays during language processing. Commonly, local structural ambiguities such as early v. late closure ambiguities are used to look at the influence of prosody on syntactic processing (Steinhauer et al., 1999; Bögels, 2011; Pauker et al., 2011). An example of such an ambiguity is (16) taken from Steinhauer, 1999 et al. (1999):

(16) *a. Peter verspricht Anna zu arbeiten # und das Buero zu putzen.*
Peter promises Anna to work # and to clean the office.
b. Peter verspricht # Anna zu entlasten # und das Buero zu putzen.
Peter promises # to support Anna # and to clean the office.
c. Peter verspricht # Anna zu arbeiten # und das Buero zu putzen.
Peter promises # Anna to work # and to clean the office.

In the German sentences, a temporary syntactic ambiguity is created at the first verb *verspricht*. It is unclear whether the verb takes a Dative object (to promise somebody (dative object) something*)* or a sentence complement (to promise something).

The second verb (*arbeiten (16 a) v. entlasten (16 b))* disambiguates the temporary syntactic ambiguity. Crucially, the absence (16 a) or presence (16 b) of a boundary (*#)* after the verb *verspricht* is supposed to cue the correct interpretation prosodically before syntactic or semantic means come into play. To look at the interaction of prosodically cued interpretation and syntactic / semantic disambiguation, a mismatch between prosodic cue (absence v. presence of the pause) and syntactic or semantic disambiguation cue can be used (16 c).

When first using such an ambiguity in German, Steinhauer et al. (1999) found a N400 / P600 pattern at the point of disambiguation (*arbeiten* v. *entlasten*). They interpreted these results as a reflection of lexical re-access to confirm the violation of the verb's argument structure (N400) and as structural revision (P600).

Bögels (2011) used a similar early v. late closure ambiguity in Dutch. She could replicate the N400 at the point of disambiguation for sentences containing a misplaced pause but no P600 was observed.

In English a similar early v. late ambiguity was used by Pauker, Itzhak, Baum and Steinhauer (2011):

(17) *a. When a bear is approaching the people # the dogs come running.*
b. When a bear is approaching # the people come running.
c. When a bear is approaching the people come running.
d. When a bear is approaching # the people # the dogs come running.

They found processing difficulty as witnessed by a biphasic N400/ P600 pattern at the second boundary for sentences like (17 d) after the first in this case misplaced misplaced boundary had already cued an intransitive interpretation of the verb *approaching*. In contrast, a missing boundary like (17c) resulted in a smaller P600. The N400 in 17 (d) is explained in terms of sentence integration difficulty. That is, the NP *the people* is prevented from getting a theta role by the two boundaries and therefore cannot be integrated in the context. The P600 in both cases is supposed to reflect revision. The authors concluded that a misplaced boundary introduces more severe processing difficulty than a missing one, which they summarised in their *Boundary Deletion Hypothesis* (BDH). It states that revision might be "easier" when no boundary is present compared to when a boundary is presented at a misleading point. That is, they hypothesise that the retrospective insertion of a missing boundary might be less effortful that undoing the positive evidence for a certain interpretation that a misplaced boundary has provided.

The BDH was recently tested by Bögels et al. (2013) who identified a number of confounds in Pauker's study. They claimed that disambiguation in Pauker's study happened on either the syntactic level via prosodic information provided by boundary on *the people* (17 d) or at the semantic level at the verb *come* (17 c).

To avoid this seeming confound they used NP (18 a. and b.) v. S (18 c. and d.) co-ordination with the disambiguation following a superfluous (18 b.) or missing PB (18 d.) always being lexical in nature.

(18) *a. The traveller followed the carrier and the guide through the mountain-like area.*

b. The traveller followed the carrier # and the guide through the mountain-like area.

c. The model kissed the designer # and the photographer took a bottle of champagne.

d. The model kissed the designer and the photographer took a bottle of champagne.

No N400 but clear a P600 effect was found for mismatching as opposed to matching conditions, for both missing (18 d v. c) and superfluous (18 b. v. a.) boundaries. However, the effect started earlier and was longer lasting for the misplaced boundary condition. The authors conclude that missing as well as misplaced prosodic information leads to processing difficulty and revision is needed. The difficulty and extent of the revision seems more pronounced for a misplaced than for a missing boundary, thus, confirming the BDH to some extent. Interestingly, these effects were only found for the first half of the experiment. During the second half of the experiment, participants seemed to have adapted to prosody being an unreliable cue in the experiment.

Kerkhofs, Vonk, Schriefers and Chwilla (2008) also looked at the influence of prosodic cues on NP v. S co-ordination in Dutch. They found an indication of processing difficulty (LAN / P600) for sentences that syntactically violated the prosodically triggered and preferred NP structure. Moreover, they also found a difference between first and second half of their experiment. During the first half a LAN effect was reported whereas for the second half a P600 was found. Again, some form of learning seems to have taken place.

Some interim conclusion can be drawn at this point. Thus far, it can be said that the absence or presence of relevant prosodic cues might trigger certain predictions about the upcoming input. When these predictions are disconfirmed, processing difficulty ensues. Also, prosodic means can change a parsing preference (such as NP co-ordination in Dutch). Moreover, the use of prosodic cues seems to be rather fluid. Listeners can adapt their reliance on prosodic cues within a single experiment.

2.5. Conclusions

In this chapter I have outlined psycholinguistic models of ambiguity resolution. These can be divided into two basic approaches. One approach assumes primacy of syntax. These are so-called serial, syntax first models of which the most prominent is the Garden-path model (Frazier, 1979, 1987). The other approach emphasises competition and the influence of certain to-be-defined constraints. They are called parallel, constraint-satisfaction models (McRae et al., 1998).

The different approaches try to account for language processing in different ways. Serial, syntax first models assume the existence of processing principles such as *Late closure* and *Minimal attachment* that are employed to guide especially syntactic processing during the initial stages (Frazier, 1979, 1987). Parallel, constraint-satisfaction models (McRae et al., 1998) on the other hand postulate that language processing is a competition-based phenomenon where different structural interpretations compete for the overall best solution of the incoming input. The winner is the structure that received most support based on constraints associated with the input. The RC attachment ambiguity has been discussed in this context.

Language processing has also been studied in the context of neurocognition of language. The models developed in this context seem to follow the assumed distinction between syntax-first (Friederici, 2002, 2011) and constraint-satisfaction (Hagoort, 2003, 2005). Different event related potentials (LAN, N400, P600) have been identified and they are used to construct models of language processing.

However, the different models are faced with a number of challenges. These centre around the variety of research questions being asked and the necessary restrictions imposed by specific research question. Also, the research methodologies employed limit the conclusions that can be drawn from studies that use them. Lastly, a wide variety of structural ambiguities is being used which might make generalisable conclusions about language processing more difficult.

I have argued that this has led to a differentiation in terms of how language processing is studied. That is, one strand of research focuses mainly on the interaction of syntax and semantics whereas the other strand seems to focus on the interaction of syntax and prosody.

The main motivation behind the current body of work is the idea that syntactic, semantic and prosodic information do not exist in isolation. This is especially true for spoken language. Therefore, the need arises to study the interaction of prosody and plausibility during syntactic structure building.

I have chosen the relative clause attachment ambiguity to study this interaction. The RC attachment ambiguity has been used to study the influence of plausibility factors and of prosodic factors on the resolution of the ambiguity in isolation. Therefore, this ambiguity presents the ideal tool to look at the interaction of prosody and plausibility during syntactic structure building.

I aimed to tackle the interaction of these factors twofold. I employed the offline method of structural priming to investigate the nature and possibly the strength of the constraints underlying RC attachment in English. The online method of ERPs is used to make inferences of the time course of the interaction.

3. Experimental methodology – structural priming and event-related potentials

In the Introduction, I have mentioned that research into relative clause attachment has utilised a number of different experimental methodologies. Studies investigating syntactic or semantic factors constraining the attachment of the RC often use reading and eye-tracking while keeping (implicit) prosodic factors constant. Studies looking into prosodic factors constraining attachment often use explicit comprehension questions or explicit (prosodic) judgements while keeping the influence of plausibility constant.

As I am studying the combined effect of overt prosody and plausibility, especially explicit comprehension might not be best suited for several reasons.

As discussed elsewhere (1.3.3.), research into the influence of prosodic factors on language processing has often used explicit comprehension or judgements. The effects of any prosodic manipulation studied with these methods are often described as subtle and small (Clifton, 2002). Moreover, as Lee and Watson (2011) point out, results from explicit comprehension specifically might not reflect syntactic processing mechanisms but rather some kind of post-sentence selection process.

Since I intended to study the interaction of prosody and plausibility in RC attachment, unbiased and possibly implicit methods are needed. Additionally, combining off-line and on-line measures seems promising, as the results should provide a differentiated picture of the underlying mechanisms and their time course. Thus, structural priming and event-related potentials were chosen.

3.1. Introduction to structural priming

Structural or syntactic priming refers to an interesting phenomenon. That observation is that people tend to re-use a certain syntactic structure of a sentence after they have just encountered that same syntactic structure. For example, someone has just heard:

1. *The teacher gave the girl the book.*

The sentence contains a verb (*gave*) that requires not just a direct object (*the book*) but also a dative object (*the girl*). That person is than presented with a picture showing an apprentice, a plumber and a spanner which she is supposed to describe. The person has at least two alternatives she can use to do so, 2 (a) and (b).

2. (a) *The apprentice handed the plumber the spanner*.
   (b) *The apprentice handed the spanner to the plumber.*

2 (a) contains the same construction as 1, that is a verb (*handed*), a direct object (*the spanner*), and a dative object (*the plumber*), whereas 2 (b) contains the verb (*handed*) a direct object *(the spanner*) followed by a prepositional phrase (*to the plumber*).

Interestingly, the person describing the picture is much more likely to use 2(a) rather than 2 (b) after first listening to 1. Thus, she has re-used the same verb + direct object and dative object construction. This repetition of a syntactic structure that was encountered in a previous sentence summarises the phenomenon of structural priming.

When priming is used as a research paradigm, participants are normally presented with a sentence such as 1. This is called the *prime*. Presentation of the prime can either involve listening to it or repeating it. Next, participants are presented with a *target*. The target can either be (amongst other things) a picture that is to be described or a sentence fragment that is to be completed, e.g. *The apprentice handed…* . The great advantage of priming as a paradigm is that people usually are not aware of re-using the same structures in consecutive trials. Thus, priming works implicitly. Contrary to comprehension questions, it is also unbiased as participants are not explicitly asked about the way the solved a possible ambiguity. Much research has shown that priming as a method is very useful (for recent reviews see Branigan, 2007; Pickering & Ferreira, 2008; Tooley & Traxler, 2010).

Interpretation of experimental results is often concerned with the nature of the mental representations that underlie syntactic processing. Because priming is an off-line measure no conclusions can be drawn about the time course of processing. However, when looking at the target that was produced it is possible to draw inferences about either the resolution of the ambiguity, the representations activated during processing and / or the weighting of the constraints applied during processing. More precisely, the structure that was produced in the target should reflect the representations activated during the previous prime trial.

Priming has been used to look at the relative clause attachment ambiguity before (e.g. Scheepers, 2003). Scheeper's (2003) could show reliable priming effects for syntactically disambiguated relative clauses (in German). Especially interesting in the current context is the finding that priming effects are prone to what Scheeper's called baseline preferences. He observed weaker priming effects for the generally preferred attachment option and stronger priming effects for the generally dispreferred attachment options. Across 2 experiments, the baseline attachment preference was taken into account. When the baseline preference showed a low attachment preference, HA primes showed stronger priming effects. When the baseline showed a high attachment preference, LA primed stronger. Thus, what seemed to drive the priming effect, was the attachment that was not in line with the observed general preference.

Thus, it can be concluded at this point, that priming is an appropriate research method to study the interaction of prosody and plausibility in RC attachment because of its implicit and unbiased nature. But it also needs to be noted that the conclusions drawn should be restricted to the possible nature of the underlying representations and not the time course of the interaction.

3.2. Introduction to event-related potentials

In order to study the time course of the interaction of prosody and plausibility in RC attachment an on-line method is needed.

One such an on-line method is the recording and analysis of electrical brain activity via electroencephalography (EEG). More specifically, a person's EEG is obtained by placing a number of electrodes on her scalp and recording the on-going electrical activity while some kind of stimulus is presented. The onset or offset of an event within the stimuli can be exactly pinpointed. By presenting many stimuli of the same kind it is possible to average across them. The process of averaging should preserve similar, recurring activity associated with processing the event while at the same time getting rid of unrelated, random activity. Since the averaged, recurring activity can be related back to the event within the stimuli, it is called Event Related Potentials (ERPs).

It is easy to understand why ERPs are an online method. Recording the electrical activity while stimuli are being presented provides researches with a continuous measure of processing. Because ERPs can be defined as a sequence of positive and negative electrical potentials, a change in the potentials can indicate how processing is affected by a certain experimental manipulation (Luck, 2005). ERPs are continuously recorded so any change in the signal is picked up immediately. Thus, the temporal resolution of on-going processing is extremely good and makes this an ideal tool to study the time course of said processing. Another advantage of ERPs is that during an experiment it is not absolutely necessary to include a behavioural task. As the change in the electrical signal is the response to the critical manipulation, researchers do not need to rely on behavioural responses (Luck, 2005). However, it has proved beneficial to include a behavioural task. ERP experiments can be more than 60 minutes long. To keep participants focused and to control whether they were actually attending what was being presented, a behavioural measure, like a judgement task or comprehension questions, can be very useful. Although ERPs are a widely used research paradigm, they have some disadvantages. As mentioned above, ERPs rely on presenting stimuli. Often a large number of stimuli is needed because changes in the signal can be very subtle and a large number of trials is required to assess them accurately. This can lead to participants' fatigue and an increase in artefacts in the signal. Another problem, which has received some more attention, is the localisation problem. Every electrical activity is characterised by the flow of a current between two sites. Thus, an electrical potential is always the amplified difference in activity between those 2 sites. In the case of EEG that difference is calculated between an active electrode and a reference electrode. More importantly, electrical potentials in the brain are very small, so any environmental electrical activity (cables etc.) will be picked up as well and distort the activity of interest. Thus, it has become common to use differential amplifiers that use three electrodes, i.e. an active, a reference and a ground electrode (Luck, 2005). The difficulty with the whole set up is that the reference can be placed virtually anywhere on the scalp or head. There are some commonly used ones like the nose or more often the left or the right mastoid or both linked together. Murray, Brunet and Michel (2008) showed convincingly how the choice of reference site effects the localisation of activity of interest. That is, the choice of reference site changes the shape, the latency and most importantly the topography of the ERPs. Suppose, one were to choose the left mastoid as the reference. Since the recorded activity is based on the difference between reference and active electrode this would magnify activity further away on the right side of the head and reduce activity closer to the reference on the left side of the head and visa versa. Similarly, linked mastoid reference amplifies activity on central sites. Thus, the choice of reference always introduces a bias (Murray et al. 2008). Consequently, only studies using the same reference could be compared safely as to the localisation of effects.

ERPs are a series of positive- and negative-going potentials. Some potentials are referred to as components and they have been distinguished on the basis of their polarity (positive vs. negative), timing (latency) and scalp distribution. The significance of a specific component or effect lays in their potential to be interpreted as a manifestation of underlying cognitive processes. I have described the most important language related components and effects in more detail in Chapter 2.

Although a number of components or effects such as (E)LAN, N400, P600, CPS have been replicated many times their functional significance is under much debate. Research into language processing draws upon many different linguistic ideas about what language is and how it is processed. This in turn is reflected in the functional interpretation of ERP components or effects as I outlined above.

This is quite probably the biggest challenge as well as the greatest opportunity in language ERP research.

3.4. Conclusion

To study the interaction of prosody and plausibility during RC attachment, two different methodological approaches will be used.

The structural priming paradigm will be employed using spoken prime sentences and to be completed target fragments. This way, the priming paradigm offers the possibility to investigate especially the representations underlying the processing of the ambiguity resolution.

ERPs will be used to study the time course of processing because of its good temporal resolution. Both methodologies complement one another; therefore a more rounded picture of the interaction should emerge.

4. Priming of relative clause attachments - Pilot study

4.1. Introduction

As detailed above (Chapter 2.), I am looking at how prosody (*pause* v. *no-pause*) and plausibility (semantic support for high v. low attachment) interact in solving the RC attachment ambiguity. We know from previous research that syntactic means, e.g. number agreement (Kaan & Swaab, 2003), plausibility means, (e.g. Cuetos & Mitchell, 1988; Carreiras & Clifton 1993; Traxler, Pickering & Clifton, 1998; van Gompel, Pickering & Traxler, 2001; van Gompel et al., 2005) and prosodic means (Clifton, 2002) can be employed to disambiguate the RC attachment between high and low. That is, prosodically, high attachment is supposed to be cued by the insertion of a pause before the RC and low attachment by no pause at this point. Additionally, plausibility information that is contained within the RC itself disambiguates attachment towards high or low.
So far, these constraints on the resolution of the RC attachment ambiguity have been investigated separately. This presents a dilemma for the present research in as much as precise hypotheses based on previous studies cannot be established easily. Thus, I decided to run an exploratory pilot study to help generate predictions about the interaction of plausibility and prosody in RC attachment.
In this pilot study, both cues were used in a completely crossed design. That is, prosodic and plausibility cues agreed in their support for attachment (*pause + plausibility for HA* v. *no pause + plausibility for LA*) or they disagreed in their support (*no pause + plausibility for HA* v. *pause + plausibility for LA*). The simplest prediction would be that they operate in an additive fashion, yielding the highest proportion of high-attachment target responses when prosody and plausibility constraints in the prime agree in their support for high-attachment (i.e., a strong prosodic boundary before a relative clause that semantically prefers to combine with NP1) and the lowest proportion of high-attachment target responses when

the two types of cues in the prime agree in their support for low-attachment (i.e., no pause before a relative clause that semantically prefers to combine with NP2). Since priming was used as research paradigm, the results might give an indication of what representations were activated during the priming stage and what representations possibly retained some residual activation during the later target processing. Thus, the aim of this pilot study was to establish what type of interaction could be expected when prosodic and semantic cues support the same or different attachment options in a RC attachment ambiguity.

4.2. Materials

Twenty-four sets of materials were created[7]. Each set consisted of pairs of sentences. Each pair consisted of a spoken prime sentence such as 1 (a-d) and a target fragment such as *The tour guide mentioned the bells of the church that… .* Each prime sentence contained a complex noun phrase (NP1-of-NP2) in direct object position with a subsequent relative clause. An example is given in (1) below.

(1)

a. *Someone shot the servant of the actress* [Pause] *who was very servile.*

b. Someone shot the servant of the actress who was very servile.

c. Someone shot the servant of the actress [Pause] *who was very famous.*

d. Someone shot the servant of the actress who was very famous.

Two versions of each prime item comprised relative clauses that would most plausibly combine with NP1 (1a and b). In the other two versions of each item the relative clause would most plausibly combine with NP2 (1 c and d). In other words, the first two versions semantically encouraged either high-attachment and the last two versions low-attachment of the relative clause. In version (1 a),

[7] The complete set of items can be found in appendix 1.

the sentence was spoken such that there was a pause before the critical relative pronoun, which is assumed to encourage high attachment of the relative clause. Version (1 d) was spoken such that there was no pause before the relative clause which is assumed to encourage low attachment of the relative clause. The items were recorded from a trained female native English speaker (L.) who is a graduate of the Performing Arts Programme of the Royal Conservatoire of Scotland. The speaker was instructed to read out the sentences using a natural intonation. Further, she was instructed to produce a pause, i.e. marked by a low boundary tone (L%) on the pre-pause word and a period of silence before the relative pronoun (Schafer, Speer, Warren & White, 2000).
Items such as (1 a) and (1 d) were read out by the speaker. Each of the two versions was recorded twice. These original recordings were -spliced at the critical point between NP2 and the relative clause and were used to create version (1 b) and (1 c).

1. a. Someone shot the servant of the actress [] who was very servile.

1. d. Someone shot the servant of the actress who was very famous.

1. b. Someone shot the servant of the actress [] who was very famous.

1. c. Someone shot the servant of the actress who was very servile.

This was done to unsure as little acoustic variability between conditions as possible. I also ensured, via post-hoc editing of the audio files, that the pause before the relative pronoun (1 a) and (1 c) was held constant at 500ms, and that no such pause occurred in the remaining conditions. To confirm that the plausibility and prosodic manipulations worked as intended, I carried out a plausibility rating study as well as acoustic analyses of the spoken materials.

4.2.1. Acoustic Analysis of the prime sentences

Apart from inserting a pause before the relative pronoun in the high-attachment prosodic disambiguation condition, I examined acoustic parameters on the noun (N2) before the relative pronoun that are also commonly assumed to mark the presence of an pause (House, 1990; Klatt, & Cooper,1975; Klatt, 1976; Wightman et al., 1992). These include the duration of last syllable of N2, as well as the fundamental frequency (F0) at the offset of N2.

There were very clear differences in the duration of the last syllable of N2. The last syllable was longer in the high-attachment prosodic disambiguation condition (Mean = 356ms, SD. = 87ms) than in the low-attachment prosodic disambiguation condition (Mean = 259ms, SD. = 85ms) with $t(46)=3.85$, $p<.001$. Correspondingly, F0 offset was significantly lower in the high-attachment prosodic disambiguation condition (Mean = 146Hz, Std. = 9.6Hz) than in the low-attachment prosodic disambiguation condition (Mean = 175Hz, Std. = 16.1Hz) with $t(46)=-7.49$, $p<.001$. These parameters, as well as the pause before the relative pronoun, should all have contributed to the perception of a pause in (1a) and (1 c).

4.2.2. Naturalness Ratings of the prime sentences

To ensure that structural priming results could not be attributed to (or masked by) potential cross-splicing artefacts, I collected naturalness ratings from a sample of 32 native English speakers who did not take part in the main experiment.

The critical stimuli were divided into four presentation files such that each file contained six items per condition (Latin square). Also included in each file were 26 filler items, recorded from the same speaker as the critical items. Half of the fillers were natural, non-edited recordings. The other half contained subtle sound manipulations such as clicks or discontinuous transitions in pitch and speech rate, mimicking acoustic impurities induced by cross-splicing. The natural and edited fillers served as comparison benchmarks for the critical items.

The rating task was carried out on a PC using DMDX. The sound files were presented via headphones in a pseudo-randomised order. Participants were instructed to focus on the acoustic features of the sound files and to judge whether the sound files were natural or edited. After listening to a sound file, they were given a five-point scale on the screen, ranging from 1 ("definitely natural") to 5 ("definitely edited"). To indicate their judgements, participants had to press a corresponding number key (1-5) on the keyboard.

Overall, the critical items (1a-d) scored a mean of 2.60 on the scale (SD = 1.63), suggesting that they were perceived as reasonably natural. In contrast, natural fillers were rated as more natural/less likely to be edited (M = 1.39; SD = 0.91) and edited fillers as less natural/more likely to be edited (M = 4.16; SD = 1.42). All three comparisons were reliable by within-subjects and between-items *t-tests* (ps < .001).

Two-way ANOVAs for the critical items revealed a main effect of *prosody* by participants only ($F_1(1,31) = 9.28$; $p < .01$; $F_2(1,23) = 2.16$; $p = .16$): the two pause conditions (1a,c) were rated as slightly less natural/more likely to be edited (M = 2.78; SD = 1.63) than the two no-pause conditions (1b,d; M = 2.42; SD = 1.62). Neither the main effect of *plausibility*, nor the *prosody* × *plausibility* interaction approached significance by either subjects or items (all ps > .4). Thus, it appears that differences in perceived naturalness across the four critical item conditions were neither very strong nor very consistent. It is also important to keep in mind that in this rating task, participants were explicitly instructed to pay attention to the acoustic features of the stimuli, whereas participants in the pilot experiment were instructed to pay attention to whether the spoken sentences made sense or not. Taken together, it seems unlikely that the results of the pilot experiment would be affected by cross-splicing artifacts in the primes.

4.2.3. Plausibility Pre-Test of the prime sentences

I collected plausibility ratings to ensure that the relative clauses in condition (1 a) semantically favoured NP1 over NP2, that the relative clauses in condition (1 d) semantically favoured NP2 over NP1,. To this end, the complex

(NP1-of-NP2-RC) noun phrases from each of the 24 items sets were reduced into simpler NP-RC combinations, resulting in four different conditions for testing (i-iv).

(i) NP1-RC1: *... a servant who was very servile.*

(ii) NP2-RC2: *... an actress who was very famous.*

(iii) NP1-RC2: *... a servant who was very famous.*

(iv) NP2-RC1: *... an actress who was very servile.*

The 24 (items) × 4 (conditions) = 96 stimuli were allotted into four lists using a Latin square (four items per condition per list). There were 48 participants so that each list was seen by eight participants. Participants were asked to rate the plausibility of each NP-RC phrase using a five-point Likert scale ranging from 1 ("not plausible at all") to 5 ("perfectly plausible"). The NP-RC phrases were preceded by the preamble "*How plausible, i.e. realistic and reasonable is ...*".

The plausibility ratings per condition are shown in Table 1. Also included in the table are results from pair-wise comparisons across the four conditions, derived from mixed-model ANOVAs treating *condition* as a fixed factor and either subjects or items as a random factor. The comparisons were based on the *Tukey* method which corrects for family-wise error.

Table 1 Mean plausibility ratings per condition

Condition	Mean (SE)		(i)	(ii)	(iii)	(iv)
NP1-RC1	4.12 (.09)	(i)	–	ns/ns	**/**	**/**
NP2-RC2	4.40 (.08)	(ii)		–	**/**	**/**
NP1-RC2	2.91 (.11)	(iii)			–	ns/ns
NP2-RC1	2.90 (.10)	(iv)				–

**: $p < .001$; ; ns: $p > .1$

RC1 relative clauses (designed to semantically favour NP1) were significantly more plausible in combination with NP1 (i) than with NP2 (iii). Conversely, RC2 relative clauses (designed to semantically favour NP2) were more plausible in combination with NP2 (ii) than with NP1 (iv). Moreover, the semantically 'matching' conditions, (i) and (ii), did not substantially differ from one another, and nor did the semantically 'mismatching' conditions, (iii) and (iv). Overall, these results confirm that the semantic manipulations worked as intended.

4.3. Priming Procedure

The experiment was carried out in DMDX (Forster & Forster, 2003), which controlled the presentation of the stimuli and audio-recorded participants' responses. Four presentation files were compiled, each containing a pseudo-random order of 24 pairs of auditory primes and written target fragments, as well as the filler materials (26 spoken sentences and 26 written sentence fragments). The four files comprised different item-condition combinations using a Latin square. Each file was seen by eight participants. There were six prime-target pairs per condition per file. Each file started with six filler trials (randomly chosen from the 52 fillers available), followed by a random sequence of 24 prime-target pairs which were separated from one another by two randomly chosen fillers.

During the experimental session, participants sat in front of a computer screen wearing a head-set with attached microphone. There were two types of trials. The first type of trial (used for the primes and spoken filler sentences) started with the prompt “LISTEN and JUDGE” on the computer screen, replaced with a fixation cross after one second. The fixation cross stayed on screen while a spoken sentence was played over the headphones. The fixation cross was then replaced with a question mark, prompting participants to indicate whether the sentence they just heard made sense or not, by saying either “yes” or “no”. The question-mark prompt stayed on screen for 4 seconds, followed by a 300 ms blank screen before the next trial was initiated. The second type of trial was used for the written target or filler sentence fragments. This type of trial started with the written prompt “COMPLETE”, which stayed on screen for 1 second, followed by the presentation of a written sentence fragment for 10 seconds. During this time, the participant’s task was to speak out a complete sentence, based on the information contained in the sentence fragment and what they thought was a sensible continuation of that sentence fragment. Audio-recordings were taken throughout the entire ten-second period, which gave participants sufficient time to complete the task. The sentence fragment was then replaced with a 300 ms blank screen before the next trial was initiated.

Since the fillers were randomly interspersed with the prime-target pairs, the sequence of “LISTEN and JUDGE” versus “COMPLETE” trials was not predictable.

4.4. Response Annotation

All target relative clause completions were transcribed and coded into one of HA (high-attachment), LA (low-attachment) or UC (unclassifiable) by a single annotator blind to condition. The critical host noun phrases within the target fragments always differed in number, and so HA and LA of the target relative clause could often be determined on the basis of number agreement between the verb within the relative clause and the relevant host noun phrase (e.g. *… the bells of the church that were/was 100 years old*). In cases where number agreement remained ambiguous, classification relied on plausibility criteria if possible (e.g. *… the bells of the church that chimed out loudly* was coded HA; *… the bells of the church that stood near the town hall* was coded LA). All other cases, including ungrammatical responses or cases where neither number agreement nor plausibility could unequivocally determine the attachment of the target relative clause, were coded as UC.

Given that the above classifications often relied on plausibility criteria, a random sample of 300 responses (39%) was independently coded by a second condition-blind annotator using the same classification scheme. Inter-annotator agreement was high (88%), with *Cohen's Kappa* indicating very good agreement at κ = .81 (± .03 SE). This confirms the validity of the first annotator's classifications which were used in the main analyses.

4.5. Subjects

Forty native English speakers (27 females) participated in the experiment in exchange for £3 or course credits. A typical session took about 25 minutes. Participants were naïve regarding the purpose of the experiment until debriefing at the close of each session.

4.6. Results

In total, 292 (30%) of the valid target responses were classified as HA, 448 (47%) as LA, and 225 (23%) as UC. Table 2 shows the target response distributions in each prime condition.

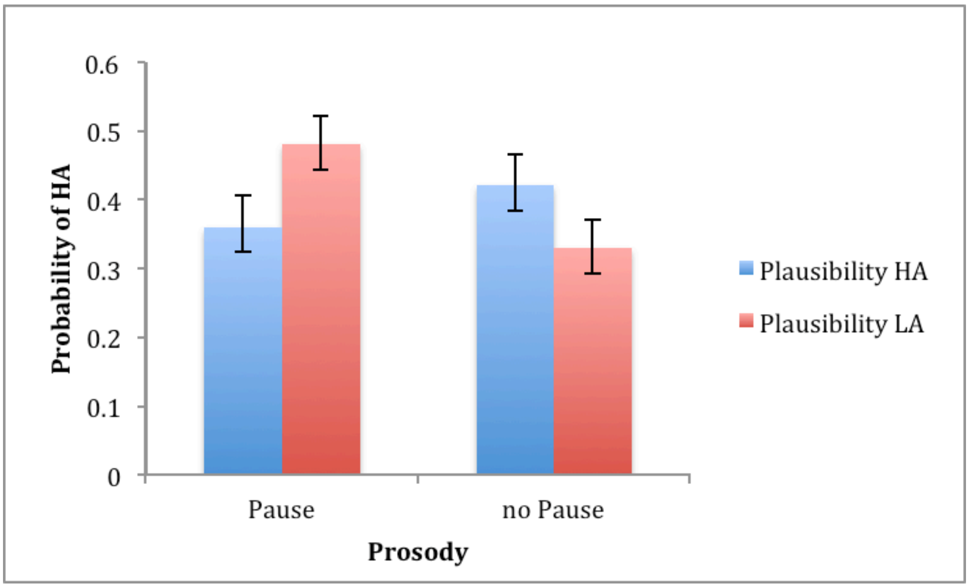
Table 2 Target response distributions per condition

Condition	Target completion HA	Target Completion LA	Target completion UC
Pause + semantic high attachment	67 27.90%	122 50.80%	51 21.30%
Pause + semantic low attachment	91 37.90%	97 40.40%	52 21.60%
No Pause + semantic high attachment	74 30.80%	106 44.20%	60 25.00%
No Pause + semantic low attachment	60 25.00%	123 51.20%	57 23.70%

Inferential analyses were based on *Generalized Estimating Equations* (GEE, e.g. Hardin & Hilbe, 2003; Hanley et al., 2003). Unlike ANOVA, this procedure allows for specifying distribution and link functions that are appropriate for categorical frequencies. The present analyses assumed a *binomial* distribution and *logit* link function (cf. Jaeger, 2008). The two predictors *prime prosody* (pause v. no pause) and *prime plausibility* (semantic support for high v. low attachment) were included as repeated-measures variables in a full-factorial 2 × 2 design using participants (GS χ^2_s), respectively items (GS χ^2_i), as reference variables for the repeated measurements. All analyses assumed an exchangeable covariance structure, and the *Generalized Score Chi Square* (GS χ^2) statistic was used for hypothesis testing.

Analyses of unclassifiable (UC) target responses in proportion to all available responses established no significant effects (all *ps* > .1); UC responses were therefore not considered further. Proportions of HA target responses out of all classifiable (HA and LA) target responses revealed no reliable main effects of either *prosody* or *plausibility* (all *ps* > .1), but a clear *prosody* × *plausibility* interaction ($\chi^2_s(1) = 8.49$; $p < .005$; $\chi^2_i(1) = 6.33$; $p < .02$). The comparison between the two no-pause conditions (1c vs. 1d) revealed a plausibility-driven priming effect, showing <u>more</u> HA target completions when the prime-RC was semantically biased towards high- (1c) than towards low-attachment (1d); 95% CIs for the simple effect: .09 ± .08 by subjects; .08 ± .08 by items. Intriguingly, the comparison between the two pause conditions (1a vs. 1b) showed a suppression effect of *plausibility*, with reliably <u>more</u> HA target completions when the prime-RC was semantically biased towards low- (1b) than towards high-attachment (1a); 95% CIs: −.13 ± .10 by subjects; −.11 ± .10 by items. Figure 1 plots the estimated marginal means (with by-subject *SE*s) per condition.

Figure 1 Estimated marginal means per condition



To conclude, neither cue to RC attachment (*prosody vs. plausibility*) resulted in a reliable main effect indicating that either prosody or plausibility takes priority.

However, we found an interaction of *prime prosody* by *prime plausibility*: without a pause before the prime-RC (prosodic support for low-attachment), plausibility cues (semantic support for either high or low attachment) showed strong priming effects. When no pause was present before the RC, I found more high attachment target completions after high attachment primes. When the prime-RC was preceded by a pause (prosodic support for high-attachment), the effect of plausibility was *suppressed*. That is the presence of a pause before the RC appeared to overrule the subsequent low attachment plausibility cue showing stronger high-attachment priming when plausibility favored low-attachment of the relative clause.

4.7. Discussion

The aim of the pilot study was to establish what type of interaction could be expected when prosodic and semantic cues support the same or different attachment options in a RC attachment ambiguity.

This pilot experiment employed a structural priming paradigm to address the question of how the two modes of disambiguation (overt prosody and plausibility) would cooperate in a fully crossed experimental design in which they would either agree or disagree in their support for high versus low attachment of the final relative clause.

Hypotheses about the interaction of prosody and plausibility could not easily be obtained from previous studies. Nonetheless, a simple hypothesis was put forward, namely that prosodic and plausibility cues could act in an additive fashion. This would result in a strong priming effect for matching as opposed to mismatching cues to RC attachment. This prediction turned out to be too simplistic as evidenced by the absence of reliable main effects of *prosody* and *plausibility* (all *p*s > .1).

On the contrary, a clear *prosody* × *plausibility* interaction ($\chi^2_s(1) = 8.49$; $p < .005$; $\chi^2_i(1) = 6.33$; $p < .02$) was found. That is, whenever the cues disagreed in their attachment, the cue associated with high attachment (pause + semantic LA and

no pause + prosodic HA) seemed to prime subsequent target completions. This interaction was unexpected. However, there might be an explanation.
To account for the observed data I am suggesting an explanation that is based on two principles. Priming of RC attachment might depend on:

> *1. The surprisal associated with a linguistic cue or constraint:*
> Assuming there is a default structural interpretation for a given structure, this principle marks a cue as surprising and thus more salient when it is not in line with the preferred default. Assuming surprisal exists, a cue that is more surprising and salient carries greater weight and will exert more influence during processing.
>
> *2. The type of structural revision necessary:*
> Assuming structural revision is caused by a mismatch between different cues, the difference in weight will determine the type of structural revision. That is, an early surprising and therefore weightier cue can overrule a later default-supporting cue; and an early default-supporting cue can be suppressed by a later surprising cue.

In short, these two principles could account for the interaction observed in the present experiment. Lets consider each in turn.
1. Surprisal associated with a linguistic cue or constraint.
It is possible that the interaction could rely on the notion of *surprisal* associated with a given disambiguation cue (Jaeger & Snider, 2008). Jaeger & Snider (2008) define surprisal as low probability of the occurrence of a certain structure. What does this mean applied to RC attachment ambiguity in English? Lets look again at the items employed in this pilot study, repeated here as (2 a-d):

(2) a. *Someone shot the servant of the actress* [Pause] *who was very servile.*

b. Someone shot the servant of the actress who was very servile.

c. Someone shot the servant of the actress [Pause] *who was very famous.*

d. Someone shot the servant of the actress who was very famous.

There is evidence that in English low attachment of the RC such as in (2 d) is preferred and, consequently, the default. Therefore, based on Jaeger and Snider (2008) and Scheepers (2003) a cue in support of the alternative high-attachment interpretation should be more surprising than a cue that is in line with the general low attachment preference. This surprisal value might make the respective HA cue more salient and effective in biasing target attachment decisions. The results show that high attachment cues are particularly effective in (2 b and c). In (2 b) overt prosodic cues support low attachment of the relative clause (no-pause condition), therefore a semantic high attachment cue might be surprising. Conversely, the plausibility cue in (2c) is in line with the low attachment default but the overt prosodic cue supports high attachment (pause before the relative clause) and should be more surprising.

One could argue, as above, the *HA (pause and semantic plausibility*) cues as such could carry surprisal value. The surprisal value is associated with more salience. Therefore, surprising cues carry more weight as a constraint on processing and exert greater influence because of it. This explanation does not, however, convincingly address the observed results. Consider (2 a). *Both* types of cues in the prime supported the less preferred high attachment interpretation (*pause, semantics HA* condition). But this did not result in any measurable increase in the number of high-attachment target responses compared to when both types of cues in the prime supported the default low-attachment interpretation such as (2 d) (*no-pause, semantics LA* condition). A post-hoc analysis confirmed this, showing 95% CIs (for the difference between these two conditions) of .03 ± .10 by subjects and .04 ± .10 by items (*p*s > .4). Hence, *surprisal of a high attachment* *cue alone* (in terms of whether it disagrees with the general low-attachment preference) may not be sufficient to explain the data. Therefore, the type of structural revision necessary might be an additional mediating factor during processing.

2. Type of structural revision necessary.

Firstly, why does revision become necessary? Put simply, during processing of a spoken sentence, a prosodic cue becomes available early on, i.e. just before the ambiguous RC. This early cue might encourage an early attachment decision, towards high (presence of the pause) or low attachment (absence of the pause). Prosodic cues are therefore likely to support *early* attachment decisions during auditory processing of the prime sentence.

Plausibility constraints are encountered later and may falsify the syntactic structure based on prosody. Thus, revision needs to take place. Indeed, the two conditions in which the two types of cues in the prime support the same relative-clause attachments (i.e., *pause + semantic HA* and *no-pause + semantic LA*) are the ones where semantically triggered structural revision is unlikely to take place. This might explain why the comparison between these two conditions showed no evidence of structural priming. Structural revisions are very likely in the remaining two conditions where prosody and plausibility support different RC-attachments (i.e., *pause + semantic LA* and *no-pause + semantic HA*). Note that both of these cue-conflict conditions elicited reliable high-attachment priming effects compared to the two no-conflict conditions.

Secondly, why should different types of revision be needed? Different types of structural revision appear to be the logical consequence of the first principle because we assumed that certain types of cues carry different weights based on surprisal.

Let's consider the cases that need structural revision:

(2) *c. Someone shot the servant of the actress* [Pause] *who was very famous.*

In the case of the prosodic high attachment cue (2 c), an early pause is a surprising and salient prosodic cue which impacts highly on processing (e.g., Speer et al., 1996; Kjelgaard & Speer, 1999; Clifton et al., 2002; Pauker et al., 2011).). Thus the presence of a pause should carry more weight than the absence of a pause which is in line with the default preference. A subsequent low attachment plausibility cue disconfirms the positive prosodic evidence. Since the pause is a strong cue, late semantic support for a different attachment might not be able to override the strong prosodic cue. As Pauker et al. (2011) argued, it might be more costly to undo the positive and highly salient evidence as encountered in a pause that was given early on. Revision needs to undue this evidence, which might be difficult to achieve.

Thus, an early surprising cue can overrule a later default-supporting cue.

In case of prosodic low attachment (2 b), the early absence of a pause in the speech signal might not be perceived as surprising because it is in line with the default preference. It might not carry the same constraint weight as a pause.

(2) *b. Someone shot the servant of the actress who was very servile.*

A subsequent high attachment plausibility cue is then associated with surprisal and leads to revision of the syntactic structure. Since there was no strong prosodic evidence early on, a surprising plausibility cue overrules the weaker prosodic evidence resulting in high attachment triggered via plausibility.

Thus, an early default cue can be overruled by a later surprising cue.

The suggested explanation for the observed priming effects can only be considered as speculative at this point. Indeed, it would be interesting to investigate a larger set of spoken materials using priming to test whether the results replicate and we look at a true effect. Additionally, an on-line method such as EEG might reveal more direct clues to the hypothesized structural revision processes than the structural priming paradigm is able to offer.

5. Materials

5.1. Introduction

In this chapter I will discuss the materials there were subsequently used for all further experiments. To be able to do that, additional considerations needed to be taken into account.

Firstly, the perceptual validation of the items used during of pilot study, returned a main effect of pause which indicated that the prosodic manipulation was perceived as very prominent. This might also be interpreted as the pause being perceived as somewhat unnatural. Thus, it was decided to include a less prominent pause. Secondly, since this new set was to be used for a priming as well as an EEG study it was thought necessary to limit the plausibility cues, i.e. the plausibility cue for high v. low attachment should only become available at the last word of the relative clause.

A new set of items was design which was subsequently validated in terms of its acoustic and perceptual properties as well as its plausibility constraints.

5.2. Items

120 sets of items were created to be used in priming as well as in an ERP experiment[8]. Normally, the number of items for behavioural experiments is comparatively small. Since these items will be used for an ERP as well a higher number was needed as detailed in Chapter 3.

Each item consisted of four spoken sentences that contained a complex noun phrase (NP1-of-NP2) in direct object position with a subsequent relative clause that could attach either high to NP1 or low to NP2. An example is given in (1) below.

[8] The complete set of items can be found in appendix 2.

(1)

a. Someone shot the servant of the actress [] who was very servile.

b. Someone shot the servant of the actress who was very servile.

c. Someone shot the servant of the actress [] who was very famous.

d. Someone shot the servant of the actress who was very famous.

In contrast to the pilot materials, the wording of the sentences of each set was identical up until the last word, thus creating a temporary syntactic ambiguity. The last word disambiguates the attachment of the relative clause. Additionally, it was decided to change the pause length from 500ms to 250ms. The pilot study was supposed to exploit prosodic features maximally. However, the validation of the material (Chapter 4.2.2.) showed that the pause conditions were judge as less natural then the no pause conditions which might have been due to the pause length. Therefore, the pause before the relative pronoun (1a and 1c) was held constant at 250ms, and no such pause occurred in the remaining conditions. The recording and post-editing was done in the same way as the pilot materials (see Chapter 4.2.)

5.2.1. Acoustic validation of the material

Different acoustic correlates are said to determine a pause. Of interest were the acoustic properties of the pre-boundary word, i.e. (1) the duration (in seconds) of the last syllable and (2) the fundamental frequency (F0 in Hz) across the last word and at the offset of NP2 as well as (3) the length of the pause. Typically some of these measures are reported in publications on prosodic features of experimental materials. However, it seems there is no consensus on what features should be reported. Thus, there is some variation of acoustic features being reported, especially with regards to the length of the pause. In the relative clause attachment literature, two studies explicitly deal with a pause (prosodic boundary) as a means to disambiguate attachment. Augurzky (2005) reports pause lengths depending on the condition between 67.4ms and 198.9ms whereas Clifton (2002) reports pause lengths from 5ms (for a weaker ip boundary) to 153ms (for a stronger IPh boundary). Pre-final lengthening is routinely described (Bögels et al., 2011; Bögels et al., 2013; Augurzky, 2005; Pauker et al. 2011; Clifton et al. 2002; Steinhauer et al., 1999) and significant differences are reported between the last syllable in pause v. no pause conditions. In the pitch domain, two approaches to reporting measures are frequent, i.e. the pitch track across a word is quantified by measuring the pitch at certain points within a word (Augurzky, 2005) or a mean pitch across the whole word is calculated (Clifton et al., 2002). Differences in the pitch trajectory or in the mean pitch are then linked to the follow-up presence or absence of a pause.

As all current materials are used in two different experimental paradigms (Priming vs. ERP) acoustic correlates of interest were measured (using Praat; Boersma, 2002) across all conditions (Tables 1 and 2). Length of the whole sentence as well as the length of phrase 1 and phrase 2 were of interest. Since the materials were used in an ERP experiment, it was necessary to check the length of the whole sentence. During an ERP experiment it is cruicial to avoid participants' blinks as this would introduce main source of artefacts in the data. Thus, the length of the sentence was checked to ensure they were short enough for participants to avoid this.

Length of the NP2 was of interest because this was the point were the prosodic manipulation came into play, i.e. pause and no pause conditions should differ significantly. Also, length of the last word was important to allow for sufficient time for possible ERP effects to manifest.

Table 1: duration in s

	1. Sentence:	2. Phrase 1:	3. Phrase 2 (excluding pause):	4. Last word:	5. NP2:	6. Last syllable NP2:	7. NP2	8. Last syllable NP2
	Someone shot the servant of the actress who was very servile / famous.	*Someone shot the servant of the*	*who was very*	*servile / famous*	*actress*	*tress*	*actress*	*tress*
N	240				120 Pause		120 no Pause	
Mean	3.842	1.913	0.618	0.647	0.602	0.395	0.489	0.287
Std Dev.	0.524	0.317	0.232	0.176	0.141	0.108	0.139	0.095
95% CI lower	3.795	1.884	0.597	0.631	0.584	0.381	0.472	0.275
95% CI upper	3.889	1.941	0.638	0.662	0.619	0.409	0.507	0.299

Pitch at the offset of NP2 was measured because pitch was expected to indicate a difference between pause and no pause conditions. The pitch before a pause should show a drop as compared to no pause.

Table 2: pitch in Hz

N	Condition	F0	Mean	Std. Dev.	95% CI lower	95% CI upper
120	Pause	Offset: actress	159.2	8.539	157.73	160.79
120	No pause	Offset: actress	165.3	8.992	163.72	166.93

Subsequent t-tests were performed on the measures of *Duration* of final syllable and *F0* at NP2 offset. On average, in the pause conditions the final syllable of NP2 is longer (*Mean* 0.395s, *SE* 0.007) than in the no pause condition (*Mean* 0.287s, *SE* 0.006). The difference was significant t(118)= 11.611, p<.001. F0 was lower in the pause conditions (*Mean* 159.2Hz, *SE* 0.779) compared to the no pause conditions (*Mean* 165.3Hz, *SE* 0.826) with t(118)= -5.357, p p<.001. Hence, the acoustic measures confirm a significant difference between pause and no pause conditions.

5.2.2. Auditory-perceptual validation of the acoustic features of the material

Acoustic measures of prosody are gradient by nature. That is why some kind of perceptual validation should be done to ensure that acoustic features translate into auditory perception in the intended way. Often the ToBI framework is used to rate and categorise the perceptional properties of boundary tones (Schafer et al., 1996; Clifton et al., 2002; Speer et al, 2003). These perceptual properties are subsequently applied to interpret differences in behavioral results. Clifton et al. (2002) for example contrasted different boundaries such as intermediate (ip) and intonational phrase boundaries and found an effect for the strength of the boundary.

Since a different speaker was used to record the present items a pretest was deemed necessary to validate the prosodic manipulation. Thus, I decided to disentangle prosodic and semantic constraints to test whether the prosodic manipulation resulted in the intended syntactic bias, i.e. the presence of the pause should trigger high attachment of the relative clause and the absence of the pause should cue low attachment of the relative clause. I opted for a forced-choice completion task based on Speer, Warren, Schafer (2003). To validate whether the pause before the RC biases the attachment of the RC, the last, i.e. the semantically disambiguating word of each sentence was cut off resulting in incomplete sentences like in Table 3:

Table 3 Materials for the perceptual pretest

Condition	
Pause	*Someone shot the servant of the actress [] who was very* ______________.
No Pause	*Someone shot the servant of the actress who was very* ______________.

The resulting incomplete utterances were divided into four lists. Each list contained 120 items, 60 in each of the two conditions, i.e. 60 incomplete sentences with a pause and 60 without a pause before the relative clause. 120 filler items were randomly interspersed with the experimental items.

4 DMDX presentation files were created. The files were counterbalanced so that each item occurred in each list once in one of the two conditions and with the two words appearing on either side of the screen. Participants listened to the sentence via headphones. They were instructed to listen to the sentences and to decide as fast and as accurately as possible which of the two words presented on the screen completes the sentence, i.e. fits best with the overall context. The subjects did not receive information that some of the sentences were ambiguous. They made the decision by pressing a button on a game pad. Buttons pressed and reaction times were recorded. 40 English native speakers took part in the experiment (29f, mean age 21,7). Participants received either £3 or course credits for participation. All were naïve in regard to the purpose of the experiment. A typical session took about 30 minutes to complete.

Overall, there were 4800 responses of which 5 were missing, i.e. a participant failed to answer or pressed a different button on the game pad. Furthermore, responses of less that 200ms and more than 5000ms were excluded from the analysis. Thus, 4570 responses entered the analysis.

Overall, responses displayed a slight high attachment preference, i.e. the overall probability of high attachment was 0.53. Descriptively, the pause v. no pause conditions differed in their probability of high attachment (0.57 for the pause v. 0.5 for the no-pause condition).

Generalised Estimating Equations modeling was employed with a *binomial* distribution and a *logit* link function. The predictor *condition* was included across participants (Generalised Chi-score *G-χ2* subjects) and items (Generalised Chi-score *G-χ2 items*). A main effect of condition was found (*G-χ2 subjects* = 13.43, $p<.001$; *G-χ2* items = 17.19, $p<.001$). Subjects interpreted the RC as attached high more often when a pause was present before the relative clause (*M subject* = .57, *SE subjects*= .015; *M items* = .56, *SE items* = .024) than not (*M subjects* = .50, *SE subjects*= .014; *M items* = .50, *SE items* = .024).

However, when looking at reaction times, the picture was not clear-cut. The GEE analysis was adjusted to a *Gamma* distribution and the predictors *condition* (pause v. no pause) and *attachment* (high v. low) were included.

There was no main effect of condition (all ps > .2). Descriptively, the decision to attach high seemed to take longer (*M subjects* = 1821ms, *SE subjects*= 98ms; *M items* = 1819ms, *SE items* = 31ms) than to attach low (*M subjects* = 1792ms, *SE subjects*= 88ms; *M items* = 1726ms, *SE items* = 31ms). Additionally, the main effect of attachment reached significance across items (*G-χ2 items*= 4.75, p= .029) but not subjects (*G-χ2 subjects* =1.302, p= .31).

An interaction of attachment*condition was found across subjects (*G-χ2 subjects* = 4.54, p= .033) but not items (*G-χ2 items* = 1.153, p= .283). That is, participants' decision to attach the target word high seemed to be facilitated by the presence of the pause (*M subjects* = 1778.77, *SE subjects*= 98.78; *M items* = 1786.33, *SE items* = 34.64) as compared to the absence of the pause (*M subjects* = 1863.28, *SE subjects*= 99.88; *M items* = 1852.03, *SE items* = 38.18). Equally, the low attachment of the target word was facilitated by the absence of the pause (*M subjects* = 1770.86, *SE subjects*= 89.15; *M items* = 1727.08, *SE items* = 39.77) as compared to the presence of the pause (*M subjects* = 1813.82, *SE subjects*= 92.98; *M items* = 1726.32, *SE items* = 35.65).

To conclude, the prosodic manipulation of the material was rather subtle. This can be seen when looking at the magnitude of the effect. The presence of a pause increased the probability of high attachment by .07, which seemed rather small.

Nonetheless, this can be taken as an indication that the pause before the relative clause supported the high attachment of the relative clause. However, this becomes only apparent in the categorical choices but less clearly in reaction times.

5.2.3. Lexico-semantic analysis of the material

Different aspects of lexico-semantic information in the host noun phrase can influence the processing of the phrase in such a way that one interpretation (implying either high- or low-attachment of the relative clause) is more salient than the other. Such factors could be e.g. differences in animacy or frequency between the host nouns in the complex noun phrase (van Gompel et al., 2005; Gilboy et al., 1995; Desmet et al., 2006). These factors were controlled for as much as possible.

I.e. concerning animacy, out of the 120 complex noun phrases 43 contained 2 animate nouns (e.g. *the wife of the warrior*), 65 contained 2 inanimate nouns (e.g. *the progress of the work*) and only 12 contained one animate and one inanimate noun (e.g. *the rep of the company* or *the website of the MSP*).

Frequency was controlled for by analysing the log frequency [9] of both nouns as shown in (Table 5):

[9] The log frequency is based on the token frequency taken from the British National Corpus. Data cited herein have been extracted from the British National Corpus Online service, managed by Oxford University Computing Services on behalf of the BNC Consortium. All rights in the texts cited are reserved.

Table 5 Log frequency of

	N	Minimum	Maximum	Mean	Std. Deviation
log_F N1	120	.000	8.79	4.79	1.92
log_F N2	120	.000	9.25	4.77	2.06

Subsequent t-tests revealed no significant differences in frequency between N1 and N2 in the log frequency $t(118)= .081$, $p=.935$.

Differences in frequency and in the number of syllables of the last word could introduce another source of error for later analyses as these words present the only lexical difference between the two versions of each item. So these were controlled for, i.e. in each item the number of syllables of the last word supporting high attachment (last word HA: *servile*) and the last word encouraging low attachment (last word LA: *famous*) were held constant. Again, the token and the log frequency of the last words were analysed (Table 6):

Table 6: Log frequency of the last words

	N	Mean	Std. Deviation
log_F last word HA	120	4.58	1.90
log_F last word LA	120	4.25	2.07

Subsequent t-tests revealed no significant differences in frequency between last word HA and last word LA, for the log frequency $t(118)= 1.248$, $p=.213$.

Hence, lexico-semantic information associated with the complex noun phrase and with the last word was controlled for as a source that might influence the attachment of the relative clause.

5.2.4. Validation of the Plausibility constraints of the material

Plausibility ratings were collected to ensure that the relative clauses in condition (1a) semantically favoured NP1 over NP2 and that the relative clauses in condition (1c) semantically favoured NP2 over NP1. To this end, the complex

(NP1-of-NP2-RC) noun phrases from each of the 120 items sets were reduced into simpler NP-RC combinations, resulting in four different conditions for testing (i-iv).

(i) NP1-RC1: *… a servant who was very servile.*
(ii) NP2-RC2: *… an actress who was very famous.*
(iii) NP1-RC2: *… a servant who was very famous.*
(iv) NP2-RC1: *… an actress who was very servile.*

The 120 (items) × 4 (conditions) = 480 stimuli were allotted into four lists using a Latin square (fourty stimuli per condition per list).

There were 32 (24 female, mean age 18.5) participants so that each list was seen by eight participants. Participants were asked to rate the plausibility of each NP-RC phrase using a five-point Likert scale ranging from 1 ("not plausible at all") to 5 ("perfectly plausible"). The NP-RC phrases were preceded by the preamble "*How plausible, i.e. realistic and reasonable is …*". Participants were compensated with 1 course credit for participation. The plausibility ratings were z-transformed to account for the between-subjects variability of using the scale.

Two-way ANOVAs with the within-subject factors *hostNP* (NP1 v. NP2) and *RC* (RC1 v. RC2) were carried out using the z-scored ratings as depended variable.

Firstly, the analyses revealed a main effect of *hostNP* by participants and items ($F_1(1,31) = 5.09$; p=.031; $F_2(1,119) = 7.19$; p = .033). Overall, host NP1 was rated as slightly less plausible (*M subjects* = .043; *SE subjects* = .019; *M items* =.043, *SE items* = .017) than the NP2 (*M subjects* = -.043; *SE subjects* = .019; *M items* =-.043, *SE items* = .017). This main effect is surprising but can be explained in the context of reducing the complex noun phrase into a single noun with an attached RC. Relational nouns such as niece, granny, father, husband occurred only in NP1 position and were reduced to a single noun with RC like

How plausible, i.e. realistic and reasonable is …

… the wife who was very beautiful.

The relational noun *wife* usually occurs with a modifying prepositional phrase (as in *the wife of the politician*) so it might have been perceived as less plausible without such a PP. More importantly, the analyses also revealed an interaction between host noun and RC ($F_1(1,31) = 463.46$; $p<.001$; $F_2(1,119) = 275.59$; $p < .001$). RC1 relative clauses designed to semantically favour NP1 were significantly more plausible in combination with NP1 (*M subjects* = .524; *SE subjects* = .027; *M items* =.524, *SE items* = .023) than with NP2 ((*M subjects* = -.530; *SE subjects* = .027; *M items* = -.530, *SE items* =.023). Conversely, RC2 relative clauses designed to semantically favour NP2 were more plausible in combination with NP2 (*M subjects* = .444; *SE subjects* = .027; *M items* =.444, *SE items* = .023) than with NP1 (*M subjects* = .-437; *SE subjects* = .027; *M items* = -.437, *SE items* = .023). Overall, this confirms that the plausibility manipulations work as intended.

5.3. Conclusion

The materials described in this chapter were to be used for three experiments. Therefore, extensive validation and pre-testing was done.

Firstly, the prosodic constraints were looked at. The analysis of the acoustic features of the stimuli revealed a difference in pre-final lengthening and pitch offset between pause and no pause conditions. The presence of the pause before the RC is associated with a longer last syllable at NP2 and a lower pitch offset. Both these features are crucial acoustic markers, however the perceptual pre-test showed that their influence on RC attachment in the current materials is rather subtle. The boundary markers in conjunction with a 250ms pause increased the likelihood of high attachment of the RC by 0.07. In contrast, the absence of the pause is associated with a shorter last syllable and a high pitch offset at NP2. The perceptual pre-test showed that the absence of the pause did not necessarily support low attachment, as the likelihood of low attachment was at chance level with 0.5.

Plausibility constraints were looked at next. During the design stage, it was ensured that there were no differences in animacy between the two nouns in the complex noun phrase. Differences in frequency were also controlled for so these possible sources of error could be excluded. The last word in each sentence has the important function of carrying the plausibility disambiguation cue. To exclude systematic differences in low level features, number of syllables and frequency of occurrence per word pair were analysed and no differences were found. Most importantly, the plausibility pre-test revealed that the disambiguation works as intended, i.e. RC1 modifies NP1 rather than NP2 and RC2 modifies NP2 rather than NP1. In sum, the materials display prosodic and plausibility constraints in the intended manne**r.**

6. Priming of relative clause attachment – main experiment 1

6.1. Introduction

The interaction of prosody and plausibility in RC attachment was first addressed in a pilot study (Chapter 4.). The aim of the pilot study was to explore in what way prosodic and semantic cues to relative clause attachment interact. A fully crossed design was used, thus the cues either agreed in their support for high (*pause and semantic HA*) or low (*no pause and semantic LA*) attachment, or disagreed in their support (*pause and semantic LA* v. *no pause and semantic HA*).
In the pilot study, no main effect of either prosody or plausibility was found, indicating that neither cue to attachment took precedence over the other. On the contrary, the analysis revealed a curious interaction of prosody and plausibility. The interaction indicated that whenever the cues disagreed in their attachment, the cue associated with high attachment (*pause + semantic LA* and *no pause + semantic HA*) seemed to prime subsequent target completions. Therefore, the aim of present experiment was foremost to investigate whether this result replicated.

6.2. Main Priming Experiment 1

6.2.1. Materials and procedure

Main priming Experiment 1 followed the same procedure as the pilot experment. The experiment was carried out in DMDX (Forster & Forster, 2003), which controlled the presentation of the stimuli and audio-recorded participants' responses.
The items were divided into 8 lists using a Latin square approach. Before items were allocated to a file they were randomly split into halves so that each file contained 60 experimental items. In the analysis, this fact was accounted for by

including the between-subjects factor "*Batch*". Not presenting all subjects with all 120 items was mainly due to two considerations. Firstly, to avoid fatiguing the subjects too much as the task was quite demanding. Secondly, to avoid long term priming effects by exposing them to the same grammatical construction over a relatively long period of time. Thus, each list contained 60 prime sentences in 4 conditions followed by 60 target fragments that were the same for all 8 lists.
60 target fragments following experimental items contained a complex noun phrase followed by a relative pronoun. The two nouns in the phrase differed in number to enable attachment disambiguation via noun-verb agreement. When pairing item and target fragment, it was controlled for that animacy features between prime and target did not overlap to avoid low-level priming, e.g.

Prime sentence:
Someone shot the servant of the actress [ ] who was very servile.
Target fragment:
The chairman explained the recommendations of the commission that_______ .

For each experimental prime as well as filler trial, subjects were instructed to perform one of two tasks. Which task to perform was prompted on the screen before each trial:

- LISTEN and JUDGE: Participants were instructed to make a global plausibility judgment of the sentence as a whole, i.e. does the sentence make sense or not.
- COMPLETE: Participants were presented with a sentence fragment that they were asked to complete in a meaningful way.

In-between the critical prime-target pairs subjects were presented with at least 2 auditory filler sentences and filler target fragments.
All participants sat in front of a monitor with a headset that included a microphone. They listened to the auditory-presented stimulus sentences while fixating a cross on the screen. Participants were asked to read the target fragment out loud for completion. Each target sentence was recorded for later analyses.

6.2.2. Response Annotation

All target relative clause completions were transcribed and coded into one of HA (high-attachment), LA (low-attachment) or UC (unclassifiable) by a single annotator blind to condition. The critical host noun phrases within the target fragments always differed in number, and so HA and LA of the target relative clause could often be determined on the basis of number agreement between the verb within the relative clause and the relevant host noun phrase (e.g. *… the <u>bells</u> of the <u>church</u> that <u>were/was</u> 100 years old*). In cases where number agreement remained ambiguous, classification relied on plausibility criteria if possible (e.g. *… the bells of the church that chimed out loudly* was coded HA; *… the bells of the church that stood near the town hall* was coded LA). All other cases, including ungrammatical responses or cases where neither number agreement nor plausibility could unequivocally determine the attachment of the target relative clause, were coded as UC.

6.2.3. Subjects

72 subjects (43 female, mean age 23.6) from the Glasgow University student body and the greater Glasgow community participated in the experiment. Each participant took part in the both acoustic threshold test (pitch and duration discrimination) as well as the listening memory span test before participating in the priming task. Overall, the three parts took no longer than 70 minutes to complete. Subjects received £7 or course credits for participation. They were naïve as to the purpose of the experiment but were debriefed after completion.

6.2.3.1. Subject variables

There were some differences between the pilot study and the present experiment. Firstly, the number of items was increased (see Chapter 5). The increase in the number of sets was deemed necessary because they were also to be used for the ERP experiment. An ERP experiment requires many more trials than standard behavioural experiments (Luck, 2005).

The new items were also recorded using a different speaker. As can be seen from Table 1, the pre-final lengthening in both sets of items was virtually the same. In contrast, the speakers realised the pitch drop at the offset of the N2 before the RC differently. Speaker L. (pilot materials) showed a greater pitch difference between pause and no pause condition than speaker N.

Table 1 Acoustic measures of materials

Materials	pre-final lengthening		F0 offset NP2	
	Pause	no Pause	Pause	no Pause
Pilot study Speaker L.	Mean = 356ms SD = 87ms	Mean = 259ms, SD = 85ms	Mean = 146Hz, SD = 9.6Hz	Mean =175Hz, SD = 16.1Hz
Experiments 1, 2 ,3 Speaker N.	Mean = 395ms SD = 108ms	Mean = 287ms SD = 95ms	Mean = 159Hz SD = 8.5Hz	Mean = 165Hz SD = 8.9Hz

Additionally, the pause length was shortened from 500ms in the pilot materials to 250ms in the present materials. The decision to shorten the pause was based on mainly one reason. That is, the pilot study explored the interaction of plausibility and prosody for the first time. Thus, the acoustic features of the materials used were designed to be very prominent to exploit them maximally. In the present materials, I opted for less prominent acoutsic features to keep the prosodic manipulation more natural.

I then compared pause length reported in previous studies in more detail (Table 2).

Table 2 Mean pause lengths

Paper	Mean pause length reported
Clifton et al. (2002)	NP1 NP2 1st break at NP1 2nd break at NP2 a: 0 ip 0ms 23ms (SD 36ms) b: ip ip 0 ms 5ms (SD 13ms) c: IPh ip 153ms (SD 53 ms) 16ms (SD 25ms)
Carreiras et al. (2004)	Pause duration (as read out of graph) ~ 250ms to 350ms
Augurzky (2005)	1st break at NP1 2nd break at NP2 HA 0.0ms LA 67.4ms HA 198.9ms LA 0.1ms
Bögels et al. (2010)	subject control: 308ms (SD 85ms) object control: 327ms (SD70ms)
Bögels et al. (2013)	NP coordination: 312 ms S-coordination: 326 ms

It can easily be seen that pause lengths varied considerably so 250ms seemed to be a good middle ground. In sum, the acoustic parameters in the present materials might be described as subtler but also more comparable to what has been reported before.

The acoustic differences in the materials naturally lead to the question whether listeners might be able to perceive a subtler boundary in the same way as a more prominent one. And, further even, might a difference in auditory perception ability impact processing of prosodically cued syntactic structures?

It has been shown that there are considerable differences in people's auditory abilities. Kidd, Watson and Gygi (2007) could show individual variation in specific abilities such as discrimination of loudness and duration, identification of highly familiar sounds of speech and non-speech and discrimination of unfamiliar simple and complex spectral and temporal patterns. Furthermore, Mo, Cole and Lee (2008) reported substantial variability in the prosodic features naïve listeners perceived in conversational speech. Most interestingly in the current context was the finding that variation in the perception of prosodic prominence and phrase boundary largely depended on the listeners.

Considering these findings, a link between auditory perception abilities and syntactic processing seems conceivable. A between-subjects factor like auditory discrimination ability should consequently be taken into account for the present experiment. It was therefore decided to record participants acoustic discrimination ability by measuring their pitch and duration discrimination ability.

Since the need arose to control for one specific between-subjects factor that might influence processing of RC attachment, it seemed sensible to control for a second factor that has been reported to account for individual differences in RC attachment. That factor was subjects' working memory span. Working memory tests measure subjects' ability to actively maintain information while processing is going on or while being distracted (Conway et al., 2005). In the context of processing the RC attachment ambiguity, differences in working memory might play a role. Indeed, it has been shown, that differences in working memory span play a role in subjects' RC attachment preferences. Traxler (2007) found a preference to attach the RC high for high span readers and low span readers preferred low attachment in an eye-tracking experiment. He concluded:

> "Cross-level interactions between working memory and attachment condition suggest that working memory capacity influences the overall difficulty readers had integrating the content of the relative clause with the preceding sentence context." (Traxler, 2007, page 1115)

In contrast, Swets, Desmet, Hambrick and Ferreira (2007) found an interaction in the opposite direction, namely high span readers preferred low attachment and low span readers preferred high attachment in an offline comprehension task. Their speculative interpretation took prosodic factors into account:

> "[…] we believe the most reasonable explanation of the relation between working memory capacity and attachment preference is that people with

> low working memory spans may tend to insert an implicit prosodic break between the complex noun phrase and the relative clause, leading to N1 attachment, whereas people with high working memory spans may tend to leave out such breaks, leading to N2 attachment." (Swets et al., 2007, page 77)

As these examples suggest, the results are far from conclusive. Apart from using different paradigms, both studies looked at overall attachment preferences and not cue conflict of some kind. Nonetheless, a working memory measure might still be helpful in investigating how subjects handle conflicting linguistic information under individual memory constraints.

To conclude, it has been shown that syntactic processing in general and processing of RC attachment specifically might be influenced by subject specific individual differences. Two areas of individual differences were taken into account for the current experiment, i.e. acoustic discrimination ability and listening memory span which were measured for each participant prior to taking part in priming study.

6.2.3.2. Pitch and duration discrimination for complex tones

I decided to use the MLP matlab toolbox (Grassi & Soranzo, 2009) to estimate participants' pitch and duration discrimination capabilities.

The discrimination procedure was based on comparing one or more standard tones that were held constant and one variable tone that changed in either pitch or duration.

I will explain the idea and use pitch discrimination as an example. In each trial a subject was presented with a number of consecutively presented tones that differed in level, i.e. one tone was higher in pitch than the others. The threshold estimation procedure varied stimuli in a way that they range from below to above the subject's discrimination threshold. For example, in one trial 3 tones were presented, two of which were the standard tones of 330Hz and one differed in pitch, e.g at 335Hz which might be above the subject's treshold. The subject correctly identified the one higher in pitch by pressing the appropriate button.

Imagine in a second trial, again the two standard tones were presented but this time the different tone has a pitch of 332Hz. The subject could not tell the difference between the tones and pressed the wrong button. In the third trial, the programme increases the pitch of the different tone until the subject can reliably indicate which tone was the different one. The pitch of that tone served as a means to calculate that subject's threshold.

To summarise, after each trial, the subject is asked to report which was the variable stimulus. To estimate the threshold, the toolbox implemented an adaptive maximum likelihood (ML) procedure detailed in Grassi and Soranzo (2009). That is, a target of correct responses (identification of the variable tone) was given. The threshold was calculated as the smallest difference between the standard tones and the variable tone a subject could identify based on the target of correct responses. The default setting of 74% for correct target responses was used.

6.2.3.3. Pitch discrimination procedure

For pitch discrimination a 330-Hz complex tone was generated that served as standard. The experiment consisted of 4 blocks of 30 trials. The subject listened to a sequence of 3 complex tones and was asked to indicate which is the highest in pitch by pressing 1, 2 or 3 on the keyboard. The first block was used as a practice block to get the subjects used to the task and it was not included in the analysis. Each subject's pitch threshold per block was recorded. The mean across the three experimental blocks was calculated.

6.2.3.4. Duration discrimination procedure

For duration discrimination a 250-ms complex tone was generated that served as standard. The experiment consists of 4 blocks of 30 trials. The subjects listened to 2 complex tones, one standard at 250ms and one variable and were asked to indicate which was the longest by pressing either 1 or 2 on the keyboard. The first block was used as a practice block to get the subjects used to the task and it was not included in the analysis. Each subject's duration threshold per block was recorded. The mean across the three experimental blocks was calculated.

6.2.3.5. Listening span test

Working memory can be thought of as a complex system that comprises a number of different subcomponents such as the ability to chunk and rehearse information to facilitate its storage. This is linked to more general abilities such as cognitive control and attention (Conway et al., 2005). Differences in this these subcomponents seem to play a role during language processing (for a review see Farmer et al, 2012). I decided to employ a memory span test that is based on Swets et al. (2007) but modified their reading span test to the suit auditory modality, i.e. I designed a listening span test. The test complied with the critical task components summarised in Conway, Kane, , Bunting, , Hambrick, , Wilhelm and Engle (2005). These componenets include some type of processing and some kind of recall of critical material. Importantly, stimuli were presented in randomised blocks of 3-6 items per block. Therefore participants were not able to predict how many items were included in a block and they could not anticipate or adapt to the upcoming working memory load. Additionally, the processing component (plausibility judgment) interfered with the recall component. This ensures that the tests measures how participants balance two demanding tasks when confronted with them in parallel.

6.2.3.6. Materials for the listening span test

Thirty-six items of the following kind were designed:

2.

(a) The letterbox had the wrong name on it but the postman did not notice.

(b) The boulder rolled down the itch and smashed into the sky.

Thirteen items were semantically correct, plausible sentences (2 a), thirteen items were semantically incorrect, implausible sentences (2 b). Each item consisted of 11-13 words which resulted in a mean length of the corresponding audio files of 3.946s (Std 0.402s). To-be-remembered (TBR) words were of high-frequency (mean log-frequency 2.36, Std 0.73), mono- or disyllabic concrete nouns. Eighteen TBR words denoted humans, e.g. farmer, boy, mother or non-human entities, e.g. horse, tree, ball. Human and non-human entities were equally distributed across semantically correct and incorrect sentences.

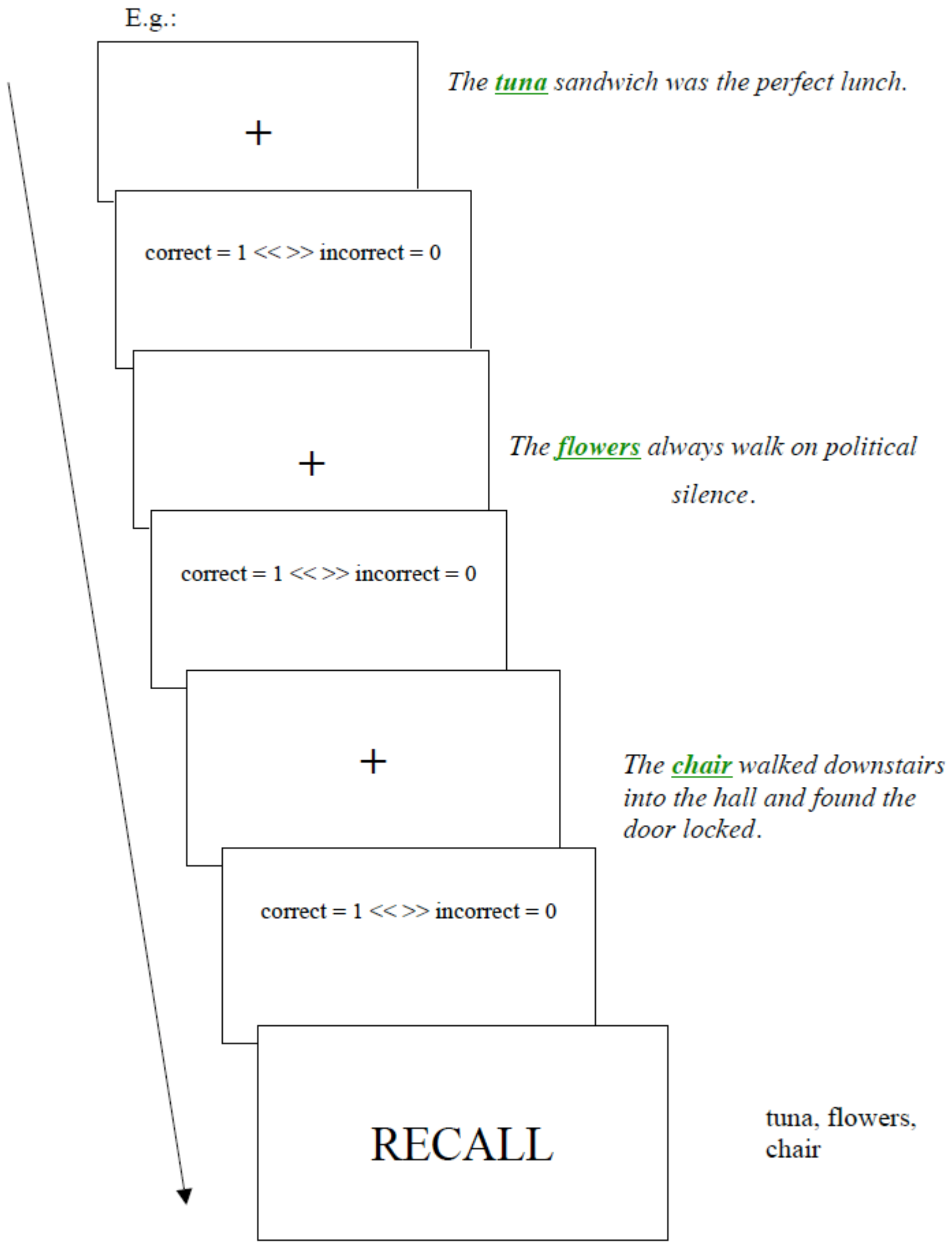
6.2.3.7. Procedure of the listening span test

The thirty-six items were randomised and divided into 8 blocks of 3 – 6 sentences. For the presentation, the blocks were randomised and always presented in the same order to minimize variability due to different presentation orders. The experiment started with a practice block of 4 sentences to get the participants accustomed to the task. The practice block was not included in the analysis.

Participants listened to the sound files via headphones and were asked to decide whether the sentence they heard made sense or not and to remember the first noun of the sentence. At the end of each block they were prompted to recall all to-be-remembered words in correct order:

E.g.:

Figure 1 Sample block for listening span test



Each block started with the prompt "Block X – Press SPACE to continue", followed by a fixation cross for 500ms. While the item was auditorily presented, a fixation cross stayed on the screen. After the sound file terminated, participants pressed either 1 for correct or 0 for incorrect and their reaction time was recorded. Subsequently, a third fixation cross appeared for 2000ms before a new trial started. At the end of each block, participants were prompted to recall all TBR words of that block in correct order of presentation. The RECALL-screen remained active for 9-18s, i.e. participants were given 3 seconds per TBR word to recall. A typical session lasted about 7 minutes. For the analysis, correct plausibility judgments were calculated (number of correct responses / 36). Number of correctly remembered words were weighed by block, i.e. number of correct words divided by the number of items in that block, then the mean across all 8 blocks was calculated. Crucially, the subject variables were recorded before the main priming experiment.

6.3. Results

6.3.1. Priming

Overall, there were 4320 valid target completions. 1252 (29%) of the valid target responses were classified as HA, 1990 (46%) as LA, and 1078 (25%) as UC. Hence, there was a general preference for low-attachment, consistent with earlier findings in English. Table 1 shows the target response distributions in each prime condition.

Table 1 Target response distributions in each prime condition

Condition	**Target completion HA**	**Target Completion LA**	**Target completion UC**
Pause + semantic high attachment	331 30.6%	482 44.6%	267 24.8%
Pause + semantic low attachment	319 29.5%	492 45.6%	269 24.9%
No Pause + semantic high attachment	312 28.9%	498 46.1%	270 25%
No Pause + semantic low attachment	290 26.9%	518 48%	272 25.1%

As can be seen, there were proportionally more HA target responses after prime sentences that were prosodically and semantically biased towards high-attachment of the relative clause, and proportionally more LA target responses after prime sentences that were prosodically and semantically biased towards low-attachment of the relative clause. However, when priming cues disagreed in their support for high or low attachment the picture was less clear.

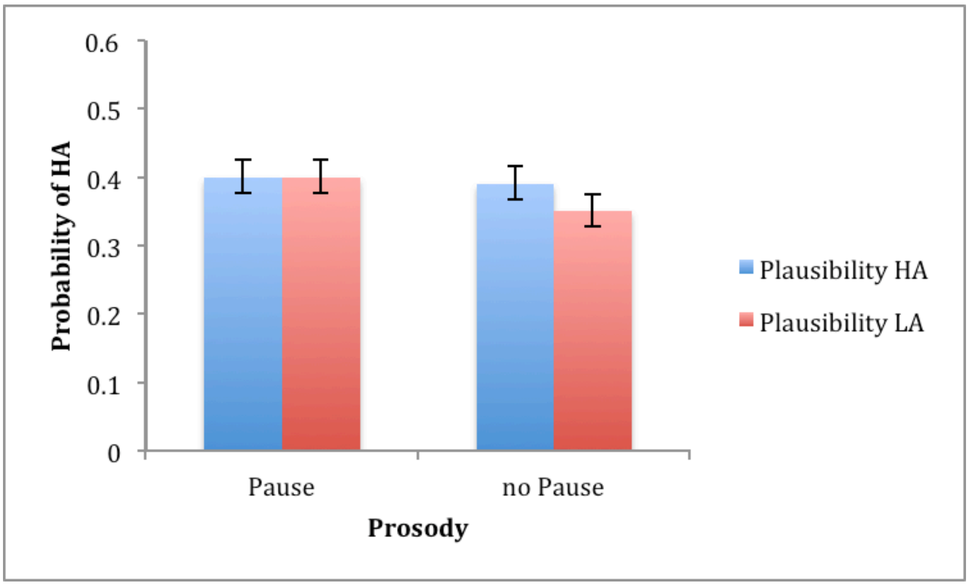
As in the pilot experiment, Generalized *Estimating Equations* (GEE, e.g. Hardin & Hilbe, 2003; Hanley et al., 2003) were carried out with a *binomial* distribution and *logit* link function (cf. Jaeger, 2008). A binomial distribution was used to account for binary nature of the data (HA v. LA or classifiable v. unclassifiable). A logit link function was used avoid empty cells.

Two within-subject predictors, i.e. *prime prosody* (pause vs. no pause), *prime semantics* (high vs. low) and the between-subjects predictor *batch* (presentation files 1-4 v. 5-8) were included as repeated-measures variables in a full-factorial 2 x 2 x 2 design using participants ($GS\chi^2_s$), respectively items ($GS\chi^2_i$), as reference variables for the repeated measurements. All analyses assumed an exchangeable covariance structure, and the *Generalized Score Chi Square* statistic was used for hypothesis testing.

The first set of analyses focused on the proportions of unclassifiable (UC) target responses out of all responses available. This analysis established no appreciable effects (all p's > .07). Thus, UC responses were not considered further.

Subsequent analyses focused on the proportions of HA target responses out of all *classifiable* target responses (HA and LA responses combined). There were no main effects of either predictor nor any interaction (all p's > .07). Figure 1 shows the estimated marginal means (with by-subject *SE*s) per condition.

Figure 1 Estimated marginal means per condition



This is somewhat surprising, as the pilot data had shown a clear *prosody* * *plausibility* interaction. However, this result is not entirely unexpected. As detailed above, there were clear differences between the materials used in the pilot study and the materials in the present experiment. Consequently, it was decided to run a cluster analysis to establish whether the overall data contained groups of participants with similar attachment behavior.

A cluster analysis (CA) is used to organise a mass of data into groups that are previously unknown. CA is a tool that classifies and groups cases based on similarity. In doing so CA does not rely on preconceived assumptions about the data. Instead clusters or groups are based solely on maximal similarity within a group while at the same time trying to maximise dissimilarity between groups (Burns & Burns, 2008). Therefore, through CA hidden patterns can be discovered based on the data alone. However, since this is a data-driven method the results might not be easy to interpret because the clusters are not defined based on theoretically meaningful criteria selected *a priori* by the experimenter. A cluster analysis will always return different clusters of observations but different methods can return very different results. So the experimenter needs to carefully evaluate which variables to choose for clustering and which clustering method to use.

6.3.2. Cluster analysis

For a cluster analysis it is important to decide which variables to use. Firstly, I chose to logit-transform the probability of high attachment in each condition for each participant. This was based on the fact that the probability of HA is used as dependent variable in the inferential analysis. I.e. I followed these steps for each participant in each condition:

1. probability HA (nr of HA / 15)
2. logit transformation:

LN((probHA + 1/120) / (1-(probHA + 1/120)))

The logit transformation was chosen to account for the categorical nature of the data. A constant of 1/120 was added to avoid empty cells.

Thus, the data were suitable for fitting a linear model. The interaction observed in the pilot study is characterised by a number of between-condition contrasts. Those include the difference between (a) pause(HA-LA), (b) no-pause(HA-LA) (c) HA(pause-no pause) and (d) LA(pause-no pause).

These differences were calcuated on a subject-to-subject basis and were subsequently used as variables in the clustering analysis.

Hierarchical clustering using the squared Euclidean distance and Ward's method was employed. Ward's method starts by assuming that every subject or case is a cluster. Cases are merged in a way that the newly formed cluster results in the least increase in the residual sum of squares. Thus, the variance within the new cluster in minimised. This method was appropriate because in contrast to other clustering methods such as *Single Linkage (SLINK), Complete Linkage (CLINK) or K means,* it is independent of the initial choice of cluster center. At the top level, the cluster analysis identified 2 different clusters which were entered into the inferential analysis separately.

6.3.2.1. Cluster 1

The cluster analysis identified 32 out of the original 72 participants to belong to cluster 1. Overall, 566 (29.5%) of the valid target responses were classified as HA, 859 (44.7%) as LA, and 495 (25.8%) as UC. Table 2 shows the target response distributions in each prime condition

Table 2 Target response distributions in each prime condition

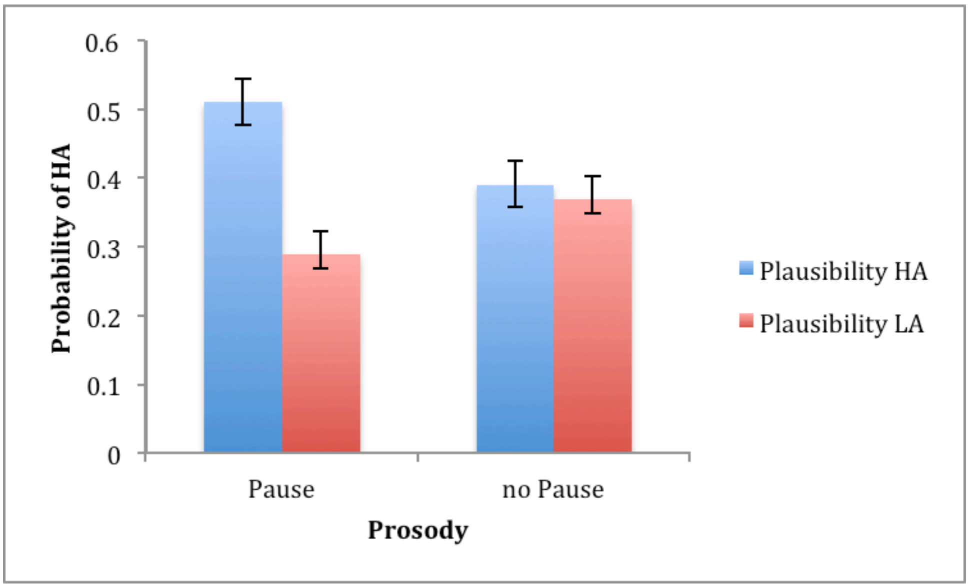
Condition	Target completion HA	Target Completion LA	Target completion UC
Pause + semantic high attachment	183 38.1%	173 36%	124 25.8%
Pause + semantic low attachment	102 21.3%	247 51.5%	131 27.3%
No Pause + semantic high attachment	142 29.6%	212 44.2%	126 26.3%
No Pause + semantic low attachment	139 29%	227 47.2%	114 23.8%

GEE analysis with a *binomial* distribution and *logit* link function was employed using *plausibility* and *prosody* as the predictors. Analyses revealed a main effect of *plausibility* (GS $\chi^2_s(1) = 16.7$; $p < .001$; GS $\chi^2_i(1) = 14.3$; $p < .001$), and a clear *prosody* × *plausibility* interaction (GS $\chi^2_s(1)=16.2$; $p <$ 001.; GS $\chi^2_i(1)=8.2$; $p = .004$).

Although a descriptive difference was present, the comparison between the two no-pause conditions (1b vs. 1d) did not replicate the plausibility-driven priming effect from the pilot study. It did not show more HA target completions when the prime-RC was semantically biased towards high- (1b) than towards low-attachment (1d); 95% CIs for the simple effect: .02 ± .05 by subjects; .03± .07 by items.

The comparison between the two pause conditions (1a vs. 1c) showed reliably more HA target completions when the prime-RC was semantically biased towards high - (1a) than towards low-attachment (1c); 95% CIs: .22 ± .05 by subjects; .19 ± .07 by items. Figure 2 plots the estimated marginal means (with by-subject *SE*s) per condition.

Figure 2 Estimated marginal means of the proportion of HA per condition



In sum, a significant interaction between *prosody* and *plausibility* was found. However, the direction of the interaction was opposite to what was expected based on the pilot study.

6.3.2.2. Cluster 2

The cluster analysis identified 40 out of the original 72 participants to belong to cluster 2.

Overall, 686 (28.6%) of the valid target responses were classified as HA, 1131 (47.1%) as LA, and 583 (24.3%) as UC. Table 3 shows the target response distributions in each prime condition.

Table 3 Target response distributions in each prime condition

Condition	**Target completion HA**	**Target Completion LA**	**Target completion UC**
Pause + semantic high attachment	148 24.7%	309 51.5%	143 23.8%
Pause + semantic low attachment	217 36.2%	245 40.8%	138 23%
No Pause + semantic high attachment	170 28.3%	286 47.7%	144 24%
No Pause + semantic low attachment	151 25.2%	291 48.5%	158 26.4%

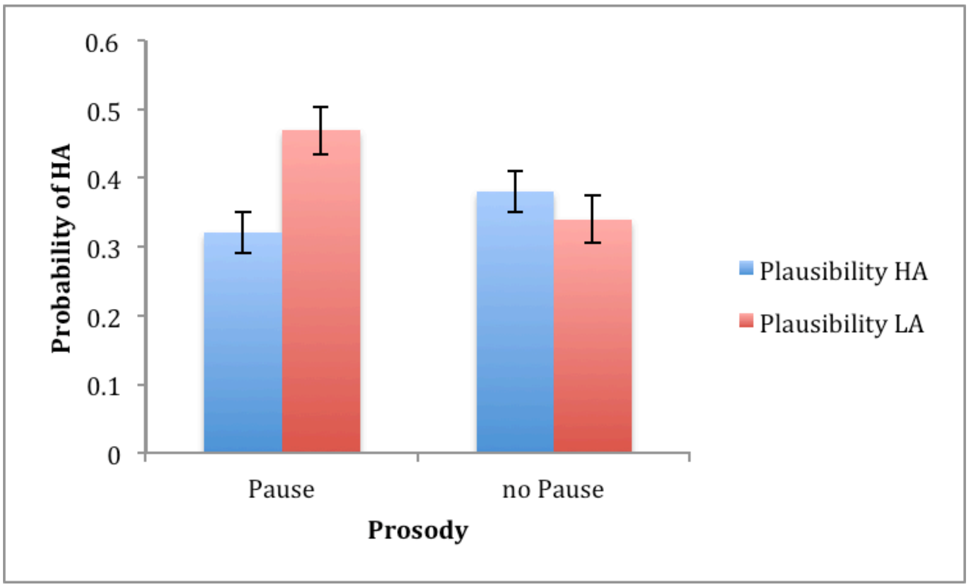
The same analyses as for cluster 1 were performed. These analyses revealed a marginal main effect of *plausibility* (GS$\chi^2_s(1)$ =4.9; p < .032; GS $\chi^2_i(1)$ =2.1; p <.139).

A clear *prosody***plausibility* interaction (GS $\chi^2_s(1)$=16.1; p < .001; GS $\chi^2_i(1)$=8.5; p = .003) was also found.

Again, a descriptive difference in the two no pause conditions was present but did not manifest itself in the GEE analysis, i.e. 95% CIs for the simple effect: .03 ± .07 by subjects; .03± .06 by items.

The comparison between the two pause conditions (1a vs. 1c) showed reliably fewer HA target completions when the prime-RC was semantically biased towards high (1a) than towards low-attachment (1c); 95% CIs: -.15 ± .05 by subjects; -.10 ± .06 by items. Figure 3 plots the estimated marginal means (with by-subject *SE*s) per condition.

Figure 3 Estimated marginal means of the proportion of HA per condition



Again a significant interaction between *prosody* and *plausibility* was found. The interaction indicated that one aspect of the pilot study could be replicated, namely I found when the prime-RC was preceded by a pause (prosodic support for high-attachment), the effect of plausibility was *suppressed*. That is, the presence of a pause before the RC overruled the subsequent low attachment plausibility cue. However, when no pause was present before the RC, I did not find more high attachment target completion after high attachment primes. I will thoroughly consider the reported results in the discussion.

To summarise, two clusters were identified that displayed opposing priming results especially in the pause conditions. The first cluster showed the strongest priming effect when both cues to high attachent (*pause + semantic support for HA*) agreed in their attachment. The second cluster showed the strongest priming effect when the cues disagreed in their attachment support (*pause + semantic support for LA*). Encouragingly, the second cluster partially replicated the results found in the pilot study. However, it is somewhat unclear what drives priming in the no-pause conditions in both clusters.

6.3.3. Subject variables

I decided to record subject-specific variables that might help explain individual differences due to the more subtle acoustic manipulation of the materials used in this study. This variables were pitch and duration discrimination ability.

Also, previous research has shown that memory span plays a role in who participants resolve the RC attachment ambiguity.

Table 4 summarises the overall descriptive analysis of the recorded subject variables pitch and duration discrimination as well as plausibility judgment and recall which make up the listening span scores overall:

Table 4 Mean pitch (Hz) and duration (ms)

	N	Mean	Std. Dev.
Pitch (Hz)	72	7.74	10.52
Duration (ms)	72	42.34	25.28
Plausibility judgment	72	.88	.14
Recall	72	.76	.18

Since I am interested in explaining the differences in attachment behavior observed in the two clusters Table 5 shows the descriptive analysis of the recorded subject variables per cluster. There were some slight descriptive differences between the two clusters:

Table 5 Mean pitch (Hz) and duration (ms) per cluster

	cluster	N	Mean	Std. Dev.
Pitch (Hz)	1	32	9.39	11.97
	2	40	6.42	9.14
Duration (ms)	1	32	43.62	22.20
	2	40	41.32	27.74
Plausibility judgment	1	32	.88	.17
	2	40	.87	.11
Recall	1	32	.73	.17
	2	40	.78	.17

For inferential analyses, the mean pitch threshold in Hz and the mean duration threshold in ms were standardised. These z-scores were summed up on a subject-to-subject basis to combine them into one variable which was taken as an indicator for the subjects' acoustic sensitivity. The plausibility judgment and recall scores were standardised and summed up as well to combine them into a single score. This z-score was taken as indicator of auditory working memory.

On average, participants (40) in cluster 2 displayed a higher sensitivity (mean_z = -.166, SE = .294) to acoustic features than participants (32) in cluster 1 (mean_z = .207, SE = .265). However, this difference was not significant $t(70)=.920$, $p=.361$.

Equally, there was an average difference in auditory working memory, i.e. cluster 2 (mean_z = .106, SE = .225) displayed a higher working memory span than cluster 1 (mean_z = .056, SE = .204). Also this difference proved non significant with $t(70)= -.160$, $p=.873$). In sum, some descriptive differences between the clusters were present but proved to be statistically non-significant.

6.4. Discussion

Main priming Experiment 1 (the present experiment) comprised a fully crossed design with the factors prosody (*pause v. no pause*) and plausibility (*semantic support for high v. low attachment*). These cues to high or low attachment either agreed in their support for high (*pause and semantic HA*) or low (*no pause and semantic LA*) attachment, or disagreed in their support (*pause and semantic LA v. no pause and semantic HA*). During the pilot study it was found that whenever the cues disagreed in their attachment support, the cue associated with high attachment (pause + semantic LA and no pause + prosodic HA) seemed to prime subsequent target completions.

The aims of the present experiment were twofold. Firstly, I intended to investigate whether results of the pilot study could be replicated. Secondly, it was pointed out that there were important acoustic differences between the materials used in the pilot and the ones in the present experiment. To be precise, the materials for the pilot study displayed a greater pitch difference at NP2 offset between pause and no pause conditions compared to the materials for the present experiment. Also, the pause following the complex noun phrase was shortened from 500ms to 250ms for the present materials. These acoustic differences could have influenced the perception of the prosodic boundary between complex noun phrase and relative clause, which might have impacted syntactic processing. Therefore participants' acoustic sensitivity, i.e. their pitch and duration discrimination ability was measured. It has also been reported (Traxler et al., 2007) that working memory capacity as measured by memory span tests interacted with participants' RC attachment preferences. Consequently, a listening span test as a measure of participants' working memory was included.

When analysing the entire dataset, no main effects or interactions were found (all p's > .07). Subsequently it was decided to run a cluster analysis. The aim of the cluster analysis was to identify unknown patterns of behaviour inherent in the data. Two clusters with opposing behavioral trends were identified. GEE analyses revealed a significant interaction of prosody and plausibility in both clusters.

The significant interaction of prosody and plausibility observed in both clusters seemed to have been driven by the match or mismatch of cues to RC attachment in the pause condition. Interestingly, the match (pause and semantic high attachment) or mismatch (pause and semantic low attachment) of cues in the prime sentences led to opposing attachment in the target completions.

Cluster 1 (32 participants), showed the strongest priming effect when both cues to RC attachment supported the same attachment, i.e. prosodic (pause) and plausibility (semantic HA) cue.

In contrast, in cluster 2 (40 participants) the strongest priming effect was observed when prosodic (pause) and plausibility (semantic HA) disagreed in their support for attachment. This replicated the finding from the pilot study. Interestingly, the clusters differed descriptively (!) in their acoustic sensitivity. The cluster that replicated the pilot results in the pause condition (cluster 2, 40 participants) showed a higher sensitivity then the one that did not replicate the pilot results (cluster 1, 32 participants).

When no pause was presented before the relative clause, the semantic HA attachment cue seemed to increase the probability of high attached in both clusters. However, the increase was not statistically significant (all p's > .25). That is, in case of cue mismatch (*no pause and semantic HA*) the semantic high attachment cue did not prime stronger then in case of cue match (no pause and semantic low attachment). Thus, neither cluster replicated the plausibility driven priming effect in the no pause condition that was observed in the pilot study.

In sum, 2 clusters were found that displayed opposing priming effects when the prosodic HA cue (pause) was present before the RC and the plausibility cue either agreed or disagreed with the prosodic HA cue. In contrast, the two clusters did not differ in terms of priming effects when the prosodic LA (no pause) cue was present at the RC and the plausibility cue either agreed or disagreed with the prosodic LA cue. More accurately, descriptively the HA cue seemed to prime in case of mismatch but the difference was not significant.

How could the difference in results between pilot and present experiment be explained?

Firstly, participants showed opposing priming effects in the present experiment as evidenced by different subgroups. They seemed to react to the prosodic HA cue (pause) in opposing ways. A reason for this could be the acoustic differences in the materials between the pilot and the present experiment. However, when comparing participants' acoustic discrimination ability no significant differences between the clusters were found although there was a descriptive difference.

Thus, at this point this measure of subject variability did not help to explain the observed results.

On the other hand, since it could be argued that the acoustic features of the materials were subtler this could have impacted participants' primability. Indeed, it has been reported before (Holyk & Pexman, 2004) that participants showed differential susceptibility to phonological priming in a lexical decision task. In two replication attempts, the authors found clear phonological priming in one and no priming in the other experiments. This was subsequently linked to participants' phonological as well as perceptual capabilities. That is, Holyk & Pexman (2004) reported that participants with higher perceptual and phonological skills showed faster reaction times to targets then participants with lower perceptual and phonological skill. Thus, this might be an indication that in this RC attachment priming experiment, individual differences in acoustic ability might be linked to primability in the pause conditions.

When discussing the results of the pilot experiment I have put forward a possible explanation for the observed interaction. In the pilot data, what seemed to prime were the high attachment cues in case of a mismatch, i.e. pause and semantic low attachment and no pause and semantic high attachment. It was argued that priming in these circumstances relied on two factors namely the *(1) surprisal* associated with the less preferred cue and *(2) the type of structural revision* required because of the conflict of cues. The explanation, as put forward, cannot account for the results observed in the present experiment. Especially the discovery of two clusters within the data that seem to display opposing priming results is associated with a certain degree of uncertainty. This is due to the fact that a cluster analysis as implemented here detects hidden patterns in the data. Whether these patterns represent a true effects can only be acertained via a direct replication. Thus, it was decided to run an exact replication of main priming Experiment 1.

7. Priming of relative clause attachment – main experiment 2

7.1. Introduction

In the pilot study a significant interaction of prosody and plausibility as cues to relative clause attachment was found. The interaction indicated that in case of cue conflict (*pause + semantic LA* and *no pause + semantic HA*) the cue associated with high attachment (pause or semantic HA) drove the priming effect. Main priming Experiment 1, however, returned some puzzling results. Firstly, the overall analysis gave no indication of priming, i.e. no main effects or interaction were found. Secondly, a more detailed cluster analysis revealed that participants seemed to respond to especially the high attachment prosodic cue (pause) in the prime sentences in different ways. Specifically two clusters of participants were found that displayed opposing priming results. Cluster 1 (32 participants) showed the strongest HA priming effect during target completion when both cues in the prime (*pause + semantic HA*) agreed in their support for high attachment as compared to when both cues disagreed in their support for high attachment (*pause + semantic LA*). Cluster 2 (40 participants), on the other hand, showed the strongest HA priming effect during target completion when both cues in the prime sentences disagreed in their support for high attachment (*pause + semantic LA*) compared to when both cues agreed in their support for high attachment (*pause + semantic HA*). Hence, cluster 2 replicated the prosody-driven priming effect found in the pilot study whereas cluster 1 did not. In contrast, no appreciable priming effects were found in both no-pause conditions, i.e. *no-pause and semantic high attachment* compared to *no-pause and semantic low attachment*.

The differences in results were partly attributed to differences in materials between the pilot study and main priming Experiment 1. The acoustic manipulation of the present materials could be described as subtler. Also, the sample of participants in priming Experiment 1 might have included a number of subjects that were less likely to be susceptible to priming.

However, in any cluster analysis, patterns of behavior can be identified which implies that a cluster analysis is very sample dependent. A test of the reliability of the different clusters would be needed to acertain whether the difference in priming behavior was due to the sample or not. It was therefore decided to replicate Experiment 1 directly.

7.2. Subject variables

For each participant in Main priming Experiment 2, the between-subject variables of listening memory span and acoustic sensitivity (pitch and duration discrimination for complex tones) were recorded. The same materials and procedures as in Experiment 1 were employed for all variables concerned (see Chapters 6.2.1. for listening memory span and 6.2.2 for pitch and duration discrimination).

7.3. Priming

7.3.1. Materials and procedure

The same materials and procedures as in Experiment 1 were employed (see Chapter 6.3.1.).

7.3.2. Response Annotation

Response annotation followed the same annotation scheme as Experiment 1 (see Chapter 6.3.2.).

7.3.3. Subjects

A second group of 72 subjects (41 female, mean age 27.3) from the Glasgow University student body and the greater Glasgow community participated in the experiment. Each participant took part in the both acoustic threshold tests (pitch and duration discrimination) as well as the listening memory span test before participating in the priming task. Overall, the three parts took no longer than 70 minutes to complete. Subjects received £7 or course credits for participation. They were naïve as to the purpose of the experiment but were debriefed after completion.

7.4. Results

7.4.1. Priming

Overall, 1314 (30.4%) of the valid target responses were classified as HA, 2320 (53.7%) as LA, and 686 (15.8%) as UC. Again, there was a general preference for low-attachment, consistent with our earlier findings and findings in other studies. Table 1 shows the target response distributions in each prime condition.

Table 1 Target response distributions in each prime condition

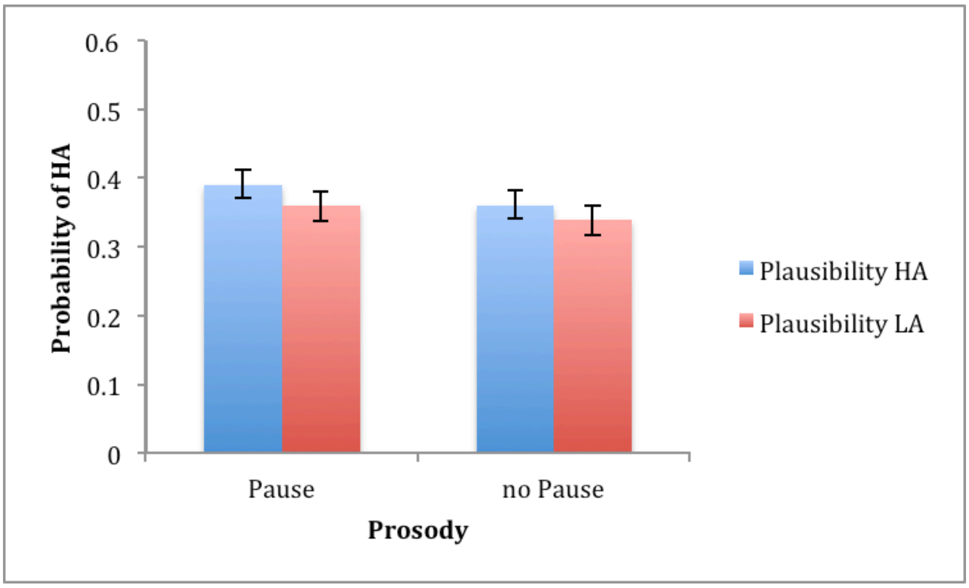
Condition	**Target completion HA**	**Target Completion LA**	**Target completion UC**
Pause + semantic high attachment	352 32.6%	558 51.7%	170 15.7%
Pause + semantic low attachment	328 30.4%	585 54.2%	167 15.5%
No Pause + semantic high attachment	323 29.9%	577 53.4%	180 16.6%
No Pause + semantic low attachment	311 28.8%	600 55.6%	169 15.6%

As in main priming Experiment 1, there were proportionally more HA target responses after prime sentences that were prosodically and semantically biased towards high-attachment of the relative clause, and proportionally more LA target responses after prime sentences that were prosodically and semantically biased towards low-attachment of the relative clause. Again, when priming cues disagreed in their support for attachment the picture was not clear.

GEEs were carried out with a *binomial* distribution and *logit* link function. The same predictors, i.e. *prime prosody* (pause vs. no pause), *prime semantics* (high vs. low) and *batch* (presentation files 1-4 v. 5-8) were included as repeated-measures variables in a full-factorial 2 x 2 x 2 design using participants (GSχ^2_s), respectively items (GS χ^2_i), as reference variables for the repeated measurements. All analyses assumed an exchangeable covariance structure, and the *Generalized Score Chi Square* statistic was used for hypothesis testing.

The first set of analyses focused on the proportions of unclassifiable (UC) target responses out of all responses available. This analysis established no appreciable effects (all p's > .08). UC responses were not considered further. Subsequent analyses focused on the proportions of HA target responses out of all *classifiable* target responses (HA and LA responses combined). There were no main effects of either predictor nor any interaction (all p's > .11). Figure 1 shows the estimated marginal means of the previously observed prosody and plausibility interaction (with by-subject *SE*s) per condition.

Figure 1 Estimated marginal means of the proportion of HA per condition



As this was not unexpected based on the results of the first main priming experiment a cluster analysis was run.

7.4.2. Cluster analysis

The same variables, data transformation and method were used as in main priming Experiment 1 (see Chapter 6.4.2). The analysis identified 2 different clusters which were entered into the inferential analysis separately.

7.4.2.1. Cluster 1

The cluster analysis identified 39 out of the original 72 participants to belong to cluster 1. Overall, 728 (31.1%) of the valid target responses were classified as HA, 1228 (52.5%) as LA, and 384 (16.5%) as UC. Table 2 shows the target response distributions in each prime condition.

Table 2 Target response distributions in each prime condition

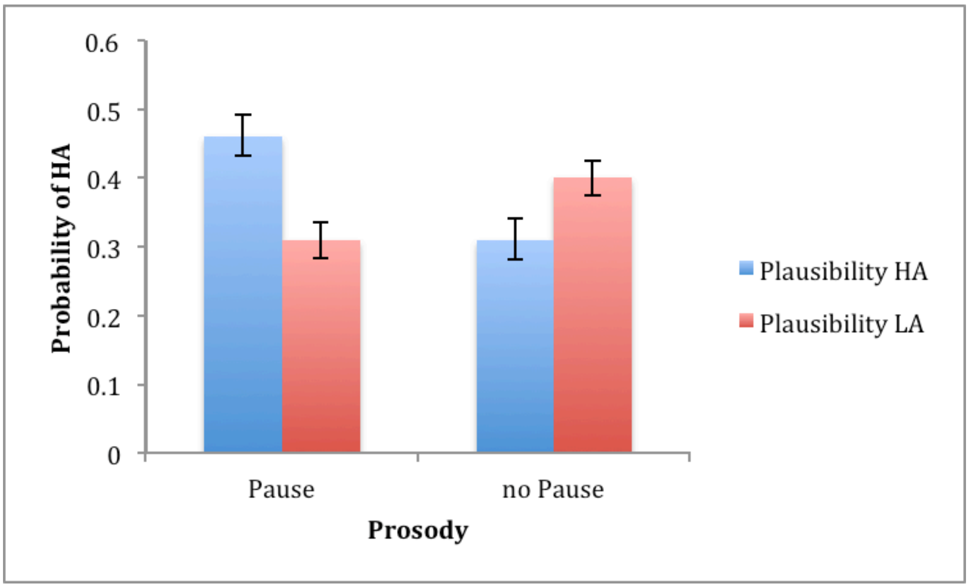
Condition	Target completion HA	Target Completion LA	Target completion UC
Pause + semantic high attachment	228 39%	265 45.3%	92 15.7%
Pause + semantic low attachment	151 25.8%	330 56.4%	104 17.8%
No Pause + semantic high attachment	149 25.5%	335 57.3%	101 17.3%
No Pause + semantic low attachment	200 34.2%	298 50.9%	87 14.8%

GEE analysis with the predictors *batch*, *prosody* and *plausibility* revealed a clear *prosody* × *plausibility* interaction (GS $\chi^2_s(1)=24.2$; p < 001.; GS $\chi^2_i(1)=18.4$; $p < .001$). No main effects or other interactions were observed.

A descriptive difference was present between the two no-pause conditions (*no-pause +semantic HA v. no-pause + semantic LA*) but in a quite unexpected direction. The comparison showed <u>more</u> HA target completions when the prime-RC was semantically biased towards low (*no-pause + semantic LA*) than towards high-attachment (*no-pause +semantic HA*); 95% CIs for the simple effect: -.09 ± .05 by subjects; -.08± .07 by items. Thus, the plausibility driven priming effect from the pilot study was not replicated.

The comparison between the two pause conditions *(pause + semantic HA v. pause + semantic LA)* showed reliably <u>more</u> HA target completions when the prime-RC was semantically biased towards high (*pause + semantic HA*) than towards low-attachment (*pause + semantic LA*); 95% CIs: .15 ± .04 by subjects; .15 ± .06 by items. Figure 2 plots the estimated marginal means (with by-subject *SE*s) per condition.

Figure 2 Estimated marginal means of the proportion of HA per condition



In sum, a significant interaction between *prosody* and *plausibility* was found. The two pause conditions replicated the result found in main priming Experiment 1. That is, in the pause conditions reliably more targets were completed with high-attached RCs when both cues in the prime supported the same attachment (*pause + semantic HA*) compared to when both cues in the prime disagreed in their support for attachment (*pause + semantic LA*). Surprisingly, in the no-pause conditions agreement for low attachment *(no pause + semantic LA*) prompted reliably more HA target completions compared to when cues did not agree in their support for attachment (*no pause + semantic HA*).

7.4.2.2. Cluster 2

The cluster analysis identified 33 out of the original 72 subjects to belong to cluster 2.

Overall, 586 (29.60%) of the valid target responses were classified as HA, 1092 (55.20%) as LA, and 302 (15.20%) as UC. Table 3 shows the target response distributions in each prime condition.

Table 3 Target response distributions in each prime condition

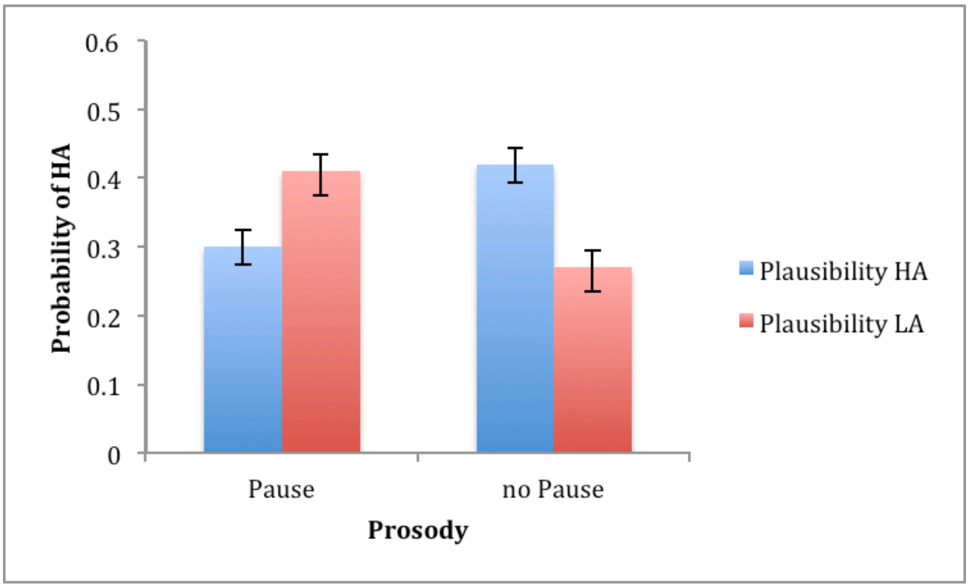
Condition	Target completion HA	Target Completion LA	Target completion UC
Pause + semantic high attachment	124 25.1%	293 59.2%	78 15.7%
Pause + semantic low attachment	177 35.8%	255 51.5%	63 12.7%
No Pause + semantic high attachment	174 35.2%	242 48.9%	79 15.9%
No Pause + semantic low attachment	111 22.4%	302 61%	82 16.5%

GEE analyses with the predictors *batch*, *prosody* and *plausibility* again revealed an interaction of *prosody* × *plausibility* (GS$\chi^2_s(1)=21.2$; $p < .000$; GS $\chi^2_i(1)=17.2$; $p < .000$).

Comparisons between the no pause conditions (*no-pause +semantic HA v. no-pause + semantic LA*) replicated the plausibility driven priming effect from the pilot study, i.e. more HA completions after semantically biased HA primes (*no-pause +semantic HA*) than after semantically biased LA primes (*no-pause + semantic LA*): 95% CIs for the simple effect: .15 ± .06 by subjects; .12 ± .06 by items.

The comparison between the two pause conditions *(pause + semantic HA v. pause + semantic LA*) showed reliably more HA target completions when the prime-RC was semantically biased towards low (*pause + semantic HA*) than towards high-attachment (*pause + semantic LA*); 95% CIs: -.12 ± .05 by subjects; -.09 ± .06 by items. Figure 3 plots the estimated marginal means (with by-subject *SE*s) per condition.

Figure 3 Estimated marginal means of the proportion of HA per condition



To sum up, a significant interaction between *prosody* and *plausibility* was found. That is, in the pause conditions, reliably more target RCs were completed attaching high when the cues disagreed in their high attachment support (*pause + semantic LA*) compared to when they agreed in attachment support (*pause + semantic HA*). This replicated the results found in the pilot as well as in main priming Experiment 1. The no-pause conditions also showed reliably more HA target completions when the cues disagreed in their support for high attachment (*no pause + semantic HA*) compared to when they agreed in attachment support (*no pause + semantic LA*). This replicated the result found in the pilot study. Priming Experiment 1 showed a descriptive result in the same direction but the inferential analysis proved non-significant.

7.4.3. Subject variables

Table 4 summerises the overall descriptive analysis of the recorded subject variables pitch and duration discrimination as well as plausibility judgment and recall which make up the listening span scores overall:

Table 4 Mean pitch (Hz) and duration (ms)

	N	Mean	Std. Dev.
Pitch (Hz)	72	6.07	7.51
Duration (ms)	72	39.17	18.19
Plausibility judgment	72	.89	.11
Recall	72	.69	.18

Again, there were descriptive differences between the two clusters, mainly in the auditory domain (Table 5):

Table 5 Mean pitch (Hz) and duration (ms) per cluster

	Cluster	N	Mean	Std. Dev.
Pitch (Hz)	1	39	6.88	9.38
	2	33	5.1	4.38
Duration (ms)	1	39	43.20	21.16
	2	33	35.14	12.88
Plausibility judgment	1	39	.89	.09
	2	33	.89	.13
Recall	1	39	.68	.18
	2	33	.69	.18

Scores that entered subsequent independent t-tests were calculated as described in main Experiment 1 (Chapter 6.4.3.). On average, participants (33) in the smaller cluster 2 displayed a higher sensitivity (mean_z = -.368, SE = .203) to acoustic features than participants (39) in the bigger cluster 1 (mean_z = .311, SE = .334). However, this difference was marginally significant at best with t(61.343)[10]=1.735, p=.08.

Additionally, there was an average difference in auditory working memory, i.e. cluster 2 (mean_z = .06, SE = .315) displayed a higher working memory span than cluster 1 (mean_z = -.051, SE = .255). The difference proved non significant with t(70)= -.279, p=.781). In sum, some descriptive differences

[10] As Levene's Test for Equality of Variances returned a significant result F=10.26, p=.002, adjusted results are reported here.

between the clusters were present but proved to be statistically non-significant.

7.5. Discussion

Main priming Experiment 2 was an exact replication of main priming experiment 1 but with a new sample of participants. In the materials, prosodic cues (pause v. no-pause) and plausibility cues (semantic HA v. LA) to relative clause attachment were used in a fully crossed design. These cues either agreed in their support for high (*pause and semantic HA*) or low (*no pause and semantic LA*) attachment, or disagreed in their support (*pause and semantic LA* v. *no pause and semantic HA*). The goal of this replication attempt was to investigate whether the clusters that were found in main priming experiment one would be found again with a new sample of participants. That is, in priming Experiment 1, two clusters of participants were found that displayed opposing attachment behaviour in the pause conditions. Cluster 2 (32 participants) showed strong priming effects when the cues mismatched (*pause + semantic LA*) whereas cluster 1 showed strong priming effects when the cues matched (*pause + semantic HA*). No reliable priming effect was found for either cluster in the no-pause conditions (*no-pause and semantic LA* v. *no pause and semantic HA*). It was hypothesised that subtler acoustic features of the prime sentences and differences in subjects' susceptibility to structural priming might be responsible for these complex results. Just as in priming Experiment 1, the overall analysis of HA target completions in the present experiment returned no significant main effects or interactions of any factors (*prime prosody*, *prime semantics* and *batch*). Encouragingly however, the cluster analysis performed on the current dataset returned two clusters that displayed opposing priming results just as in main priming Experiment 1. Specifically, cluster 1 of the present experiment (39 participants) showed an interaction of *prosody* and *plausibility.* When comparing the two pause conditions, results revealed more high attachments when the relative clause was disambiguated towards high rather than low attachment. When comparing the two no-pause conditions this cluster showed a reversal of the expected effect, i.e. more high attachments when the relative clause was supporting low attachment rather than high attachment.

Cluster 2 in the present experiment (33 participants) showed an interaction of *prosody* and *plausibility*. That is, in the pause conditions, a strong priming effect was found for the mismatch of cues (*pause and semantic LA*) as compared to the match of cues (*pause and semantic HA*). In the no-pause conditions again, a strong priming effect was found for the mismatch of cues (*no-pause and semantic HA*) as compared to the match of cues (*no-pause and semantic LA*). However, as before, a significant difference between the cluster in terms of acoustic discrimination ability or listening memory span was not found, although, again, a descriptive difference was present with cluster 2 (33 participants) showing a greater acoustic sensitivity as measured by pitch and duration discrimination than cluster 1 (39 participants).

In sum, the presence of two different clusters with opposing priming effects that was found in main priming Experiment 1 could be replicated in main priming Experiment 2. Importantly, the clusters shared the results observed in the *pause conditions*. That is, clusters 1 (exp. 1 with 32 and exp. 2 with 39 participants) showed s strong priming effect when the cues supported the same attachment (*pause + semantic HA*). In contrast, clusters 2 (exp. 1 with 40 and exp. 2 with 33 participants) showed strong priming effect when the cues supported different attachments (*pause + semantic LA*). However, there were some important differences as well, especially in the no-pause conditions. Whereas no reliable priming effects could be reported for both clusters in the first main priming experiment, there were clear and somewhat surprising priming effects in the no-pause conditions in the second main priming experiment. That is, cluster 1 (39 participants) showed most high attachment of target RCs when both cues supported low attachment *(no-pause + semantic LA)*, cluster 2 showed strong priming effects in the mismatching case (*no-pause + semantic HA*).

Let's return to the goals of this experiment at this point. One purpose of the present experiment was to investigate whether certain sources of error could account for apparent differences in results between the pilot experiment and main priming Experiment 1. I will consider each in turn.

1. *Differences in the acoustic features between pilot (24 items) and present (120 items) materials*: It was hypothesised that the differences in acoustic features of the prime sentences might be responsible for the observed differences. Remember that the materials in the pilot study marked the pause by a greater pitch drop at NP2 offset and a 500ms pause before the RC. The present materials comprised a smaller pitch drop at NP2 offset and a shorter, 250ms pause. The results of priming Experiment 2 implies that the difference in acoustic features did impact the perception of the pause and possibly even processing of the RC attachment ambiguity. The discovery of two clusters that seemed to react in opposing ways to the presence of the pause was an indication for this. There was some descriptive evidence that this might be linked to participants' acoustic discrimination abilities.

2. *Individual differences in susceptibility towards priming:* It was proposed (Holyk & Pexman, 2004) that there might be individual differences between participants in terms of primability and that this could be linked to their perceptual and phonological abilities. Intriguingly, the cluster that replicated the initial pilot results also displayed higher acoustic sensitivity than the cluster that did not. This is in line with the idea that it takes either a very strong prosodic manipulation or very sensitive participants to produce those results. The difference in acoustic sensitivity between these clusters was close to significant (p=.08). This might be taken as a first, tentative indication of possible individual differences in how particippants responded to the priming cues. I will discuss the implications of this in more detail in Chapter 8.

7.6. Conclusion

So far, I have reported experimental results of 3 priming experiments. To get an overview of all results up until this point and to be able to appreciate the commonalities and differences between the single experiments, it seems beneficial to summarise them in Table 5.

Table 5 probability of high attachment in the target RC per experiment, cluster and condition (standard error in brackets)

		Condition			
Cluster (Number of participants)	Experiment	Pause + semantic HA	Pause + semantic LA	No Pause + semantic HA	No Pause + semantic LA
1 (32)	Main prime 1	.51 (.035)	.29 (.032)	.39 (.032)	.37 (.022)
1 (39)	Main prime 2	.46 (.032)	.31 (.028)	.31 (.026)	.40 (.026)
2 (40)	Main prime 1	.32 (.03)	.47 (.034)	.38 (.03)	.34 (.035)
2 (33)	Main prime 2	.30 (.024)	.42 (.024)	.41 (.024)	.27 (.035)
40	Pilot	.36 (.045)	.48 (.036)	.42 (.041)	.33 (.037)

The aim of this thesis was to investigate the interaction of prosodic cues (pause v. no-pause) and plausibility cues (semantic HA v. LA) in resolving the RC attachment ambiguity.

As can be seen from the table, there was a pattern of results that seemed to reoccur. Evidence was found that priming seemed to rely on disagreement of prosodic and semantic cues. That is, whenever the cues disagreed in their attachment, the cue associated with high attachment (*pause + semantic LA* and *no pause + semantic HA*) seemed to prime subsequent target completions. This pattern was found in the pilot study, in main priming Experiment 2 (cluster 2 with 33 participants) and partly in main priming Experiment 1 (cluster 2 with

40 participants). On the other hand, a substantial proportion of participants did not show this priming result. On the contrary, the only reliable and replicated result indicated that the strongest priming effect occurred when prosodic (pause) and plausibility (semantic HA) cue both supported high attachment. This mixed set of results made any theoretical interpretation and grounding difficult. However, a possible solution was inherent in the fact that main priming Experiment 2 was a direct replication of main priming Experiment 1. Therefore, before any evaluation of the results as to their possible theoretical implications was attempted, it was decided to combine both experiments to increase the power of the statistic tests.

8. Priming experiments 1 and 2

8.1. Introduction

In both main priming experiments that were run (Chapters 6. and 7.), the overall analysis did not return consistent significant main effects or interactions. Subsequent cluster analyses identified 2 subsets of participants in each experiment that seemed to display opposing priming behavior. In the first main priming experiment, the clusters responded to the match (*pause + semantic HA*) or mismatch (*pause + semantic LA*) of the prosodic and semantic cues differently in that cluster 1 showed more HA target completions after matching (*pause + semantic HA*) and cluster 2 after mismatching cues (*pause + semantic LA*). No priming effects were found for no-pause conditions in this experiment. In the second main priming experiment, again the clusters showed the same priming effects in the pause conditions, i.e. cluster 1 had more HA target completions after matching cues (*pause + semantic HA*) and cluster 2 had more HA target completions after mismatching cues (*pause + semantic LA*). Both clusters differed in their priming results in the no-pause conditions. When comparing the two no-pause conditions, cluster 1 showed an unexpected effect, i.e. more high attachments for matching low attachment cues (*no-pause + semantic LA*). In cluster 2, in comparison, the mismatch (*no-pause + semantic HA*) primed strongest.

However, as the number of cases in the clusters differed substantially between experiments as can be seen in Table 1.

Table 1 Number of subjects per cluster and experiment

Experiment	N Cluster 1 Acoustically less sensitive	N Cluster 2 Acoustically more sensitive
1	39	33
2	32	40

The number of participants per cluster influences the statistical power of the test used. Since the clusters differed in number of participants in each cluster it was decided to combine both experiments to increase the statistical power. Combing both datasets was made possible by the fact that both experiments employed the same materials and procedures, i.e. priming Experiment 2 was an exact replication of priming Experiment 1. Thus, cluster 1 of Experiment 1 and cluster 1 of Experiment 2 were combined to form 1 overall cluster of 71 participants. Also, cluster 2 of Experiment 1 and cluster 2 of Experiment 2 were combined to form 1 overall cluster of 73 participants.

8.2. Results

8.2.1. Overall Analysis

Across both experiments, 2566 (29.7%) of the valid target responses were classified as HA, 4310 (49.9%) as LA, and 1764 (20.5%) as UC. As was to be expected, the general low attachment preference persisted. Table 2 shows the target response distributions in each prime condition.

Table 2 Target response distributions in each prime condition

Condition	**Target Completion HA**	**Target Completion LA**	**Target completion UC**
Pause + semantic high attachment	683 31.6%	1040 48.1%	437 20.2%
Pause + semantic low attachment	647 30%	1077 49.9%	436 20.1%
No Pause + semantic high attachment	635 29.4%	1075 49.8%	450 20.9%
No Pause + semantic low attachment	601 27.8%	1118 51.8%	441 20.4%

As can be seen, there were more HA target responses after prime sentences that prosodically and semantically supported high-attachment than low attachment. Conversely, I found marginally more LA target responses after prime sentences that were prosodically and semantically biased towards low rather than high attachment. There seemed to be hardly a difference in target completions when priming cues disagreed in their support for attachment.

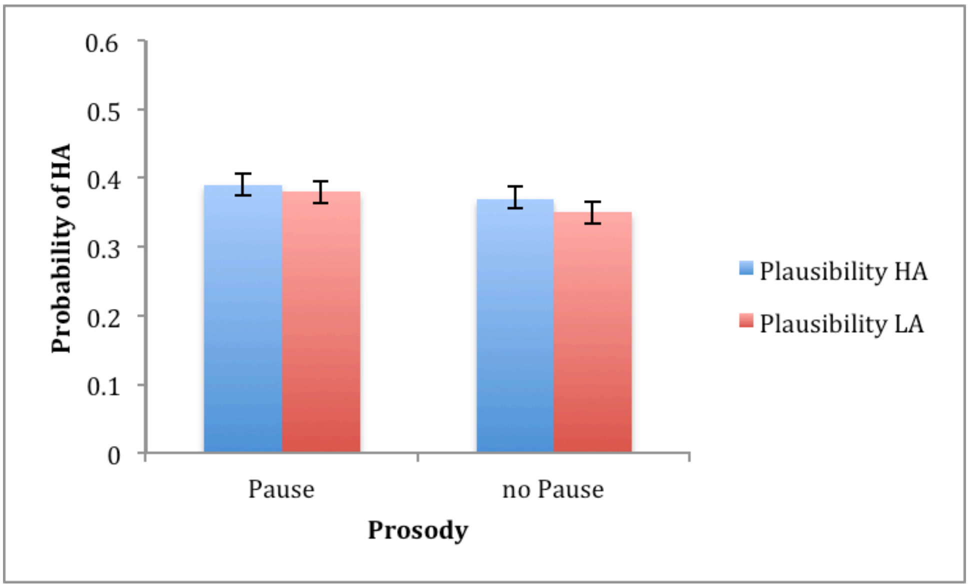
Before analysing the proportion of high attachment target completions, unclassifiable target completions were examined. As with previous analyses, GEE modeling with a *binomial* distribution and a *logit* link function was employed using the predictors *prosody* (pause vs. no pause), *plausibility* (high vs. low), *batch* (presentation files 1-4 v. 5-8) and *experiment* (priming Experiment 1 vs. 2). There was a significant main effect of *experiment* across subjects and items (GS$\chi^2_s(1)=40.39$; $p < .001$; GS$\chi^2_i(1)=50.68$; $p < .001$). This indicated that there were more unclassifiable target completions in Experiment 1 than 2. This difference can be attributed to a small number of subjects in the first experiments that had a high number of UC target completions. As I did not find any other effects (ps > .10) UCs were excluded from further analysis.

The next set of analyses focus on the proportion of high attachment target completions out of all classifiable target completions (HA and LA responses combined). A main effect of prosody was found ($\chi^2_s(1)=5.36$; $p = .021$; $\chi^2_i(1)=4.08$; $p = .043$), indicating more HA target completions when a pause was present (*Mean subjects* = .39, *SE subjects* = .014; *Mean items* = .39, *SE items* = .020) that when it was absent (*Mean subjects* = .36, *SE subjects* = .013; *Mean items* = .36, *SE items* = .020). This is interesting insofar as it was the pause that seemed to play a major role in all previous cluster analyses.

Additionally, main effects of *plausibility* (($\chi^2_s(1)=3.37$; p = .070; $\chi^2_i(1)=4.57$; p = .032) and *experiment* ($\chi^2_s(1)=.887$; p = .346; $\chi^2_i(1)=5.11$; p = .024) were also found across items but both did not reach significance across subjects. That is, the main effect of *experiment* indicated that Experiment 1 had a slightly higher overall probability of high attachment (*Mean subjects* = .38, *SE subjects* = .019; *Mean items* = .39, *SE items* = .020) than experiment 2 (*Mean subjects* = .36, *SE subjects* = .016; *Mean items* = .36, *SE items* = .020). The main effect of *experiment* might be linked to the fact that there was a considerable difference between experiments in terms of *unclassifiable target completions*. Overall, Experiment 1 comprised 25% unclassifiables whereas Experiment 2 had 15.8% unclassifiables. The higher number of classifiable target completions in experiment 2 seemed to have impacted on low attachment target completions thereby decreasing the probability of high attachment in the second experiment overall. The main effect of *plausibility* indicated that the plausibility constraints in Experiment 1 led to a slightly higher overall probability of high attachment (*Mean subjects* = .38, *SE subjects* = .014; *Mean items* = .39, *SE items* = .020) than Experiment 2 (*Mean subjects* = .36, *SE subjects* = .016; *Mean items* = .36, *SE items* = .020). The main effect of *plausibility* could also possibly be linked to the higher number of low attachment target completions based on fewer unclassifiables. Additionally, individual differences in participants' susceptibility to priming could come into play. This seemed to be supported by the fact that the main effects did not reach significance across subject.

As these effects are not consistent across subjects and items I will refrain from interpreting them any further. No other significant main effects or interactions were found (all ps > .08). Figure 1 plots the estimated marginal means (with by-subject *SE*s) per condition.

Figure 1 Estimated marginal means of the proportion of HA per condition



In sum, the overall analysis for both main priming experiments seemed to corroborate the results reported for the single experiments. Particularly, the prosodic manipulation (pause) seemed to impact processing of the RC attachment ambiguity reliably as evidenced by the consistent main effect. Other factors such as plausibility or experiment did not result in generalisable effects. Since previous analyses hinted at the presence of substantial individual difference, subsequent analyses focused on the clusters identified previously.

8.2.2. Cluster 1

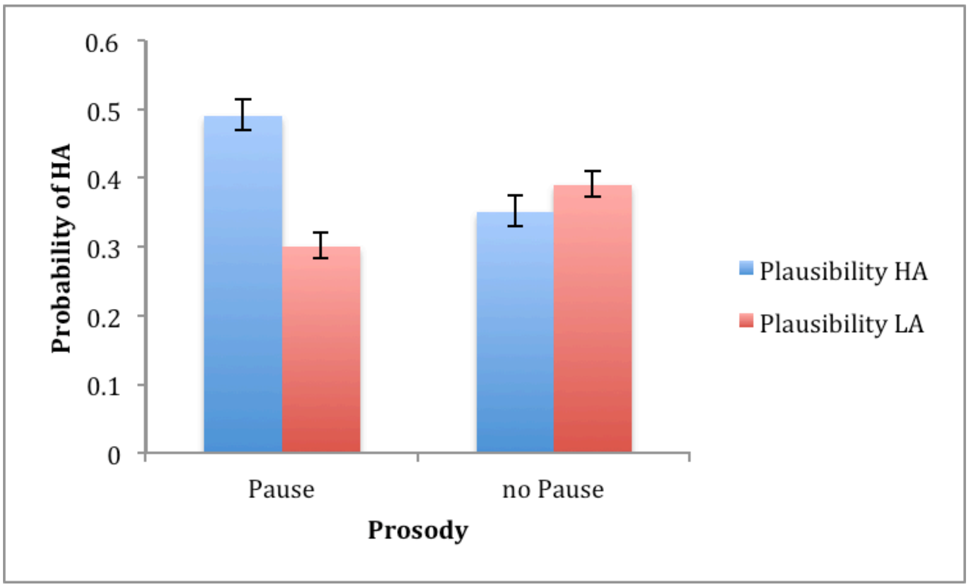
Cluster 1 consisted of 71 participants out of 144 original participants. Overall, 1294 (30.4%) of the valid target responses were classified as HA, 2087 (49%) as LA, and 879 (20.6%) as UC. Table 3 shows the target response distributions in each prime condition.

Table 3 Target response distributions in each prime condition

Condition	**Target Completion HA**	**Target Completion LA**	**Target completion UC**
Pause + semantic high attachment	411 38.6%	438 41.1%	216 20.3%
Pause + semantic low attachment	253 23.8%	577 54.2%	235 22.1%
No Pause + semantic high attachment	291 27.3%	547 51.4%	227 21.3%
No Pause + semantic low attachment	339 31.8%	525 49.3%	201 18.9%

GEE analysis with the predictors *batch*, *prosody*, *plausibility* and *experiment* revealed a main effect of *plausibility* (GSχ^2_s(1)=20.5; p < 001.; GS χ^2_i(1)=13.7; p < .001) and a clear *prosody* × *plausibility* interaction (GS χ^2_s(1)=38.4; p < 001.; GS χ^2_i(1)=21.9; p < .001). The interaction seemed to be driven by the comparison of the two pause conditions (*pause + semantic HA v. pause + semantic LA*). The comparison showed reliably more HA target completions when the prime-RC was semantically biased towards high - (*pause + semantic HA*) than towards low-attachment (*pause + semantic LA*); 95% CIs: .18 ± .04 by subjects; .17 ± .05 by items. When looking at the two no-pause conditions (*no-pause +semantic HA v. no-pause + semantic LA*), the comparison showed more HA target completions when the prime-RC was semantically biased towards low - (*no-pause + semantic LA*) than towards high-attachment (*no-pause +semantic HA*). 95% CIs for the simple effect (.04 ± .04 by subjects; .02 ± .05 by items) underlined the observation that it was unclear what was driving priming in the no-pause conditions. However, it seemed save to conclude that the plausibility driven priming effect found in the no-pause conditions in pilot study was not replicated in this cluster. Figure 2 plots the estimated marginal means (with by-subject *SE*s) per condition.

Figure 2 Estimated marginal means (with by-subject *SE*s) per condition



To conclude, this cluster (71 participants) showed a strong priming effect in one pause conditions. More specifically, the agreement of prosodic and plausibility cues in their support for high attachment (*pause + semantic HA*) led to the strongest priming effect. What determines priming in the no-pause conditions based in the results remained unclear.

8.2.3. Cluster 2

Cluster 2 consisted of 73 participants out of 144 original participants. Overall, 1272 (29%) of the valid target responses were classified as HA, 2223 (50.8%) as LA, and 885 (20.2%) as UC. Table 4 shows the target response distributions in each prime condition.

Table 4 Target response distributions in each prime condition

Condition	Target completion HA	Target Completion LA	Target completion UC
Pause + semantic	272	602	221
high attachment	24.8%	55%	20.2%
Pause + semantic low	394	500	201
attachment	36%	45.7%	18.4%
No Pause + semantic	344	528	223
high attachment	31.4%	48.2%	20.3%
No Pause + semantic	262	593	240
low attachment	23.9%	54.2%	21.9%

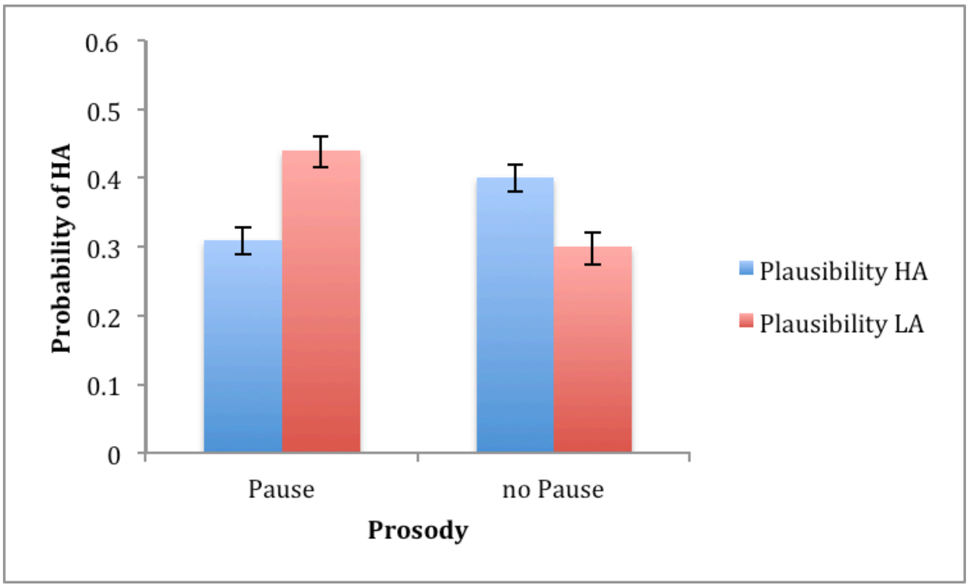
GEE analyses with the predictors *batch*, *prosody*, *plausibility* and *experiment* revealed a clear *prosody* × *plausibility* interaction (GS $\chi^2_s(1)=38.0$; $p <$ 001.; GS $\chi^2_i(1)=21.9$; $p < .001$).

The comparison between the two no-pause conditions (*no-pause +semantic HA v. no-pause + semantic LA*) showed reliably more HA target completions when the prime-RC was semantically biased towards high (*no-pause +semantic HA*) than towards low-attachment (*no-pause + semantic LA*); 95% CIs: .09 ± .05 by subjects; .07 ± .04 by items. Thus, the plausibility driven priming effect found in the pilot study has been replicated when clusters 2 of both priming experiments are combined.

The comparison between the two pause (*pause + semantic HA v. pause + semantic LA*) conditions showed reliably more HA target completions when the prime-RC was semantically biased towards low - (*pause + semantic LA*) than towards high -attachment (*pause + semantic HA*); 95% CIs: -.13 ± .04 by subjects; -.09 ± .04 by items. Again, the results confirmed the effect found in the pilot study.

Figure 3 plots the estimated marginal means (with by-subject *SE*s) per condition.

Figure Estimated marginal means of the proportion of HA per condition



To conclude, this cluster showed a strong priming effect in one of the two pause conditions. This effect seemed to be driven by the mismatch between the prosodic high attachment cue (pause) and the plausibility cue (semantic support for LA). In the two no-pause conditions, priming seemed to rely on the high attachment plausibility cue.

8.2.4. Subject variables

Table 5 summarises the overall descriptive analysis of the recorded subject variables pitch and duration threshold as well as plausibility judgment and recall which made up the listening span scores overall:

Table 5 Descriptive analysis of pitch and duration threshold, plausibility judgment and recall

	N	Mean	Std. Dev.
Pitch (Hz)	144	6.91	9.02
Duration (ms)	144	68.88	21.74
Plausibility judgment	144	0.89	0.13
Recall	144	0.73	0.18

As expected from the previous analyses, there were descriptive differences between the two clusters, mainly in the auditory domain (Table 6):

Table 6 Descriptive analysis of pitch and duration threshold, plausibility judgment and recall per cluster

	cluster	N	Mean	Std. Dev.
Pitch (Hz)	1	71	8.14	10.68
	2	73	5.76	6.76
Duration (ms)	1	71	43.41	21.68
	2	73	38.23	20.31
Plausibility judgment	1	71	0.89	0.13
	2	73	0.88	0.12
Recall	1	71	0.71	0.18
	2	73	0.74	0.18

Firstly, the raw pitch (in Hz), duration (in ms), plausibility judgment and recall scores were z-scored. The z-scores were then summed up to be entered in to the subsequent independent t-tests as combined variables. One variable expressed acoustic discrimination ability. The other variable expressed auditory listening span.

On average, participants (73) in cluster 2 displayed a higher sensitivity (mean_z = -.257, SE = .185) to acoustic features than participants (71) in cluster 1 (mean_z = .264, SE = .218). However, this difference was not significant t(137.66)[11]=1.826, p=.07. There was a slight average difference in auditory working memory, i.e. cluster 2 (mean_z = .08, SE = .187) displayed a higher working memory span than cluster 1 (mean_z = -.0026, SE = .166). The difference proved non significant with t(142)= -.352, p=.725).

To conclude, 2 subject-specific variables were recorded to determine whether the factors acoustic discrimination ability and auditory listening span might account for differences found in participants priming results.

Auditory listening span does not seem to play any appreciable role. But, I found

[11] As Levene's Test for Equality of Variances returned a significant result F=6.72, p=.010, adjusted results are reported here.

descriptive differences in acoustic discrimination ability although these were not supported by the inferential analysis (p=.07). The way acoustic discrimination ability was measured included solely pitch and duration discrimination of complex single tones. However, prosodic phenomena are acoustically more complex as they involve changes in pitch and duration over the course of at least a sentence. Thus, the present measure might not have captured the whole spectrum of individual acoustic abilities and a more complex test, such as the Montreal Battery of Evaluation of *Amusia* (Peretz, Champod & Hyde, 2003) should be employed in future follow-up studies.

Nonetheless, the observed descriptive differences hint at the fact that auditory discrimination ability might play a role.

Thus, I decided to run a correlation analysis to firstly explore whether there is a relationship between participants' membership in a cluster and participants' acoustic discrimination ability. Secondly, I was interested in the size a possible effect might have. There was a significant relationship between the cluster a participant belonged to and their acoustic discrimination ability, r=.152, p(one-tailed)=.035. That is, participants in cluster 1 tended to have a higher score for acoustic discrimination ability than participants in cluster 2. This meant that participants in cluster 2 could reliably discriminate smaller pitch and duration differences between tones than participants in cluster 1. Additionally, Pearson's correlation coefficient of r=.152 indicated that the effect is rather small.

8.3. Discussion

Two major priming experiments were run independently of one another. Each experiment used a fully crossed design of prosodic (*pause v. no-pause)* and plausibility (*semantic support for HA v. LA*) cues to RC attachment. That is, prosodic and plausibility cues either agreed in their support for attachment (*pause + plausibility for HA* v. *no pause + plausibility for LA*) or they disagreed in their support (*no pause + plausibility for HA* v. *pause + plausibility for LA*). Since both experiments shared the set-up, procedures and items, it was decided to combine the independent datasets to increase the power of the statistical analyses.

The overall analysis of the combined datasets returned a main effect of prosody. The presence of a pause seemed to increase the probability of high attachment target completions independent of other factors. This result was not unexpected. When validating the prosodic manipulations in a pretest (Chapter 5.2.2.) it was the pause that increased the probability of high attachment in a forced-choice completion task. Therefore, an effect of pause could be expected.

However, an interesting observation occurred while analysing the first main priming experiment. That is, 2 clusters of participants were identified that showed dissimilar priming results. One cluster was found that seemed to react to disagreement of prosodic and semantic cues. That is, whenever the cues disagreed in their attachment, the cue associated with high attachment (*pause + semantic LA* and *no pause + semantic HA*) seemed to prime subsequent target completions. A second cluster, on the other hand, showed the strongest priming effect when prosodic (*pause*) and plausibility (*semantic HA*) cues both supported high attachment. It remained unclear what drove priming in this cluster in the no-pause conditions. Identical clusters were reported in the second priming experiment. Comparable clusters were found in each individual experiment, it seemed reasonable to combine them, thereby increasing the statistical power of the inferential analyses.

Results for cluster 1 (71 participants) indicated that priming of high attachment target completions in the pause conditions was driven by combined support for high attachment (*pause + semantic HA).* In contrast, the result in the no-pause conditions gave no consistent indication of priming. Results for cluster 2 (73 participants) revealed that priming was driven by the high attachment cue in case of cue mismatch (*pause + semantic LA* and *no-pause + semantic HA*). Interestingly, analyses of the subject-specific variables implied that acoustic sensitivity as measured by pitch and duration discrimination might play a role in interpreting these diverging priming results. However, the variable measuring listening memory span did not show significant differences between the clusters.

From previous research it is known that plausibility, (e.g. Cuetos & Mitchell, 1988; Carreiras & Clifton 1993; Traxler, Pickering & Clifton, 1998; van Gompel, Pickering & Traxler, 2001; van Gompel et al., 2005) and prosodic means (Clifton, 2002) can reliably and independently disambiguate RC attachment between high and low.

In this thesis, I am looking at the interaction of prosody and plausibility in solving the RC attachment ambiguity. The pilot study was run under the preliminary assumption that prosody and plausibility could operate in an additive fashion. That is, whenever the cues agree in their support for either high or low attachment (*pause + semantic HA* and *no-pause + semantic LA*) strong priming effects would be expected. The results reported for pilot study as well as for main priming Experiment 1 and 2 and their combined analyses clearly falsify this assumption. I did not find the highest number of high attachment target completions for matching HA cues (*pause + semantic HA*) and the lowest number of HA target completions for matching LA cues (*no-pause + semantic LA*). The results of the pilot study suggested that whenever the cues disagreed in their attachment, the cue associated with high attachment (*pause + semantic LA* and *no pause + prosodic HA*) seemed to prime subsequent target completions. I put forward two principles that might account for the observed results of the pilot study. Priming of RC attachment might depend on the *surprisal associated with a linguistic cue or constraint,* and the *type of structural revision* necessary. However, these two principles as they stand, cannot account for the observed results of main priming Experiment 1 and 2 and their combined analyses. At least a third principle needs to be added at this stage. As I will show below, the third principle might be a prerequisite for the already suggested principles. Priming of RC attachment might depend on:

1. *The prominence of the prosodic high attachment cue (pause):*

   Assuming acoustic features of a pause (such as pitch drop at and lengthening of the pre-pause word, duration of silence) were being perceived as more or less prominent, this difference in prominence might impact processing of a prosodically supported syntactic structure. This prominence is dependent on the prosodic cue itself and on the perceivers' acoustic sensitivity.

2. The surprisal associated with a linguistic cue or constraint:

Assuming there is a default structural interpretation for a given structure, this principle marks a cue as surprising and thus more salient when it is not in line with the preferred default. Assuming surprisal exists, a cue that is more surprising and salient carries greater weight and will exert more influence during processing.

3. The type of structural revision necessary:

Assuming structural revision is caused by a mismatch between different cues, the difference in weight will determine the type of structural revision. That is, an early surprising and therefore weightier cue can overrule a later default-supporting cue; and an early default-supporting cue can be suppressed by a later surprising cue.

These three principles could help to interpret the differential priming results observed in the previous experiments. However, the first principle might influence the degree to which the subsequent two principles might exert influence during processing. I will therefore consider prosodic salience first and will then look at what the difference in salience might mean in terms of syntactic processing in different listeners.

1. The prominence of the prosodic high attachment cue (pause):

Prosodic boundaries and pauses are marked by acoustic features such as pitch drop at the end of the pre-boundary word, lengthening of the pre-boundary syllable and the duration of the pause itself. Differences in these parameters lead to differences in strength of the pause present in the speech signal as suggested for example by the differentiation between strong boundaries such as IPhs and weaker intermediate boundaries such as ips (Pierrehumbert, 1980). Stronger boundaries such as IPhs have been shown to impact processing of ambiguous phrases more, especially with regards to the surrounding prosodic context (e.g.

Clifton et al., 2002; Carlson et al., 2009). In addition, it has been shown previously that there is considerable between-subject variation in the perception of a prosodic boundary (Mo & Lee, 2008). It has also been reported that there are considerable differences in people's auditory abilities (Kidd et al., 2007).

Based on these considerations one could argue that prosodic boundaries are perceived to higher or lesser degree depending on acoustic differences in the speech signal itself as well as the listeners acoustic sensitivity. If acoustic information is to be taken into account during syntactic processing it should have been perceived as important or in this case surprising. The importance or surprisal value of acoustic information depends on its acoustic noticeability, which in turn could depend on the individuals' perception. Taking this reasoning as a starting point, subsequent processing and the concepts of surprisal and revision could be interpreted in this context. The question that needs to be answered in terms of the resolution of the RC attachment ambiguity could therefore be: Do listeners perceive the prosodic high attachment cue (pause) as "prominent (surprising) enough" to influence subsequent processing?

2. The surprisal associated with a linguistic cue or constraint:

I have argued before (Chapter 4.7.) that the interaction of *prosody and plausibility* could rely on the surprisal associated with a given disambiguation cue (Jaeger & Snider, 2008). Jaeger and Snider (2008) define surprisal as low probability of the occurrence of a certain structure. The data of the priming results detailed above suggest an additional assumption. To evaluate the probability of occurrence the parser should have perceived the occurrence of a certain linguistic cue such as a pause and thus, be surprised by it. If a surprising low probability cue such as a pause or semantic HA information has been perceived, it gains added weight during processing, thus impacting syntactic processing. The surprisal of a linguistic cue, especially in terms of prosody, might not just be determined by its low probability of occurrence but also by its noticeability. If the pause is distinct enough it is noted as surprising and therefore carries weight in terms of subsequent processing. If it is not distinct enough its surprisal value is lessened and the cue would carry less weight for subsequent processing.

3. The type of structural revision necessary:

Structural revision might become necessary during processing of a spoken sentence because a prosodic cue (*pause* v. *no-pause*) becomes available early on, i.e. just before the ambiguous RC. This cue is subsequently confirmed or disconfirmed by plausibility information (*semantic HA* v. *LA*). In cases of a mismatch revision of the current structure might be needed. I have argued that different types of structural revision processes might be found based on the weight or surprisal value of the temporally preceding cue in comparison the weight or surprisal value of the current cue. To reiterate, *an early surprising cue such as a pause can overrule a later default-supporting cue such as semantic LA. Also, an early default-supporting cue such as a no-pause can be overruled by a later surprising cue such as semantic HA.* What does this mean applied to RC attachment ambiguity with respect to the different clusters found? Lets look again at the items employed, repeated here as (1 a-d):

(1) a. *Someone shot the servant of the actress* [Pause] *who was very servile.*

b. Someone shot the servant of the actress who was very servile.

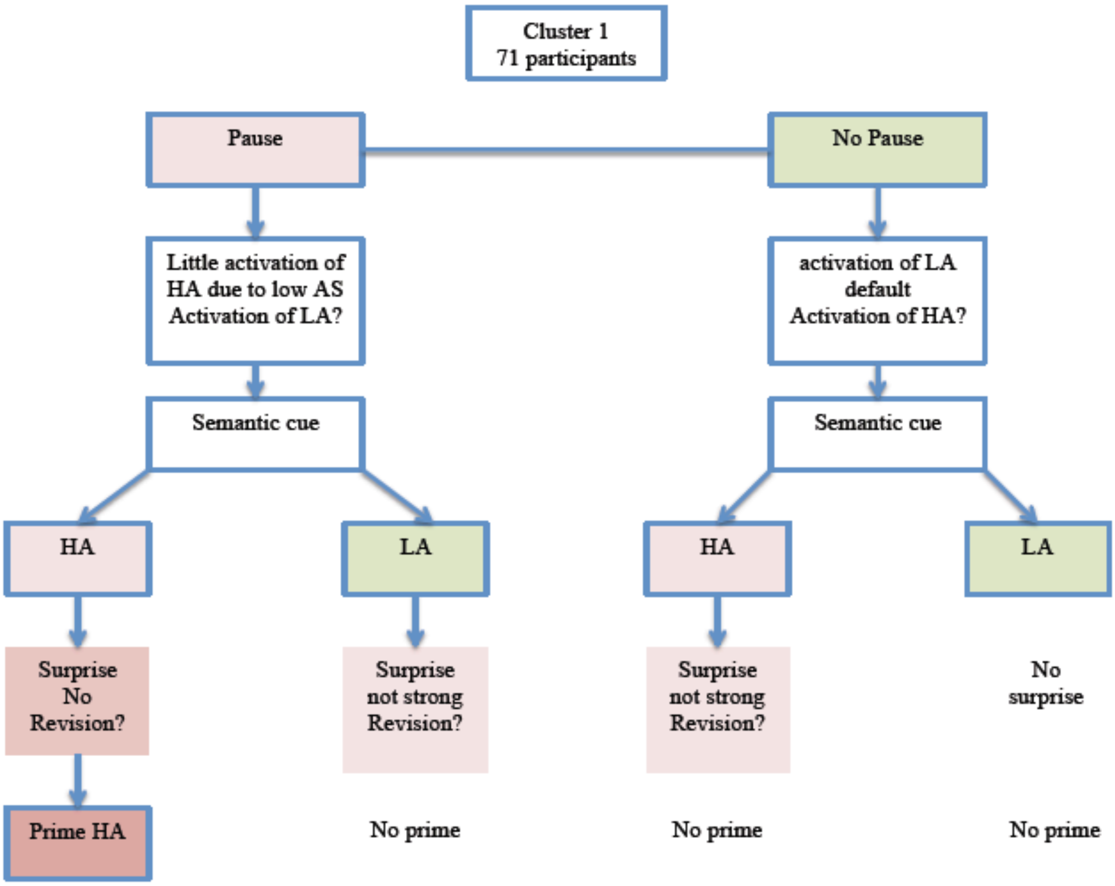
c. Someone shot the servant of the actress [Pause] *who was very famous.*

d. Someone shot the servant of the actress who was very famous.

1. *Cluster 1 (71 subjects) prosody and plausibility interaction*

There is some evidence that low attachment of the RC such as in (1 d) is preferred in English and, consequently, might be the default relative clause attachment (e.g. Cuetos & Mitchell, 1988). Therefore, based on Jaeger and Snider (2008) a cue in support of the alternative high-attachment interpretation (*pause* v. *semantic HA*) should be more surprising than a cue that is in line with the general low attachment preference. This surprisal value might make the respective HA cue more salient and effective in biasing target attachment decisions. However, subtler acoustic features of the prosodic high attachment cue might make it more or less difficult for listeners to note the occurrence of this cue. One might even go as far to assume that low acoustic sensitivity might lead to less reliance on prosodic cues for this cluster in general. Figure 1 visualises the influence of the proposed principles during priming for cluster 1.

Figure 1



Assume a pause was present in the speech signal (1 a and c) and has not been perceived as prominent and surprising enough due to lower acoustic sensitivity (AS). This should consequently add little weight to the syntactic HA option. It is equally possible that the alternative syntactic LA structure might be pursued alongside the high attachment structure. When a high attachment semantic cue is encountered (1 a) it is perceived as surprising because it is not in line with the preferred LA option. Thus, extra weight is added to the HA option which subsequently primes due to overall surprisal. Indeed, cluster 1 showed the strongest priming effect when both HA cues matched (*pause + semantic HA*).

In contrast, again assume a pause was present in the speech signal but has not been perceived as prominent and surprising enough due to lower acoustic sensitivity. Consequently little weight is added to the syntactic HA option and the alternative syntactic LA structure might be pursued alongside the high attachment structure. When a low attachment semantic cue is encountered (1 c) it is perceived as in line with the preferred LA option and might create a mismatch. Thus, revision might occur. Priming now should depend on the type of structural revision that needs to take place, i.e. *an early surprising cue such as a pause could overrule a later default-supporting cue such as low attachment semantic information.* However, I did not find any priming effects that did indicate revision and subsequent priming of the surprising pause cue. This could be due to the pause being perceived as not surprising enough. Therefore there are 2 possible explanations for not finding priming effects. One explanation could be that revision did occur but took longer because the prosodic HA cue did not add enough surprisal value to sufficiently differentiate between alternative syntactic HA v. LA structures. This would argue for some sort of ongoing competition during revision. Thus no priming occurred. The second explanation could be that revision occurred but was resolved towards low attachment because the prosodic HA cue was not strong and surprising enough. Therefore no priming occurred. Thus, *an early surprising cue such as a pause could overrule a later default-supporting cue such as semantic LA* ***only if*** *the pause was perceived as sufficiently surprising.*

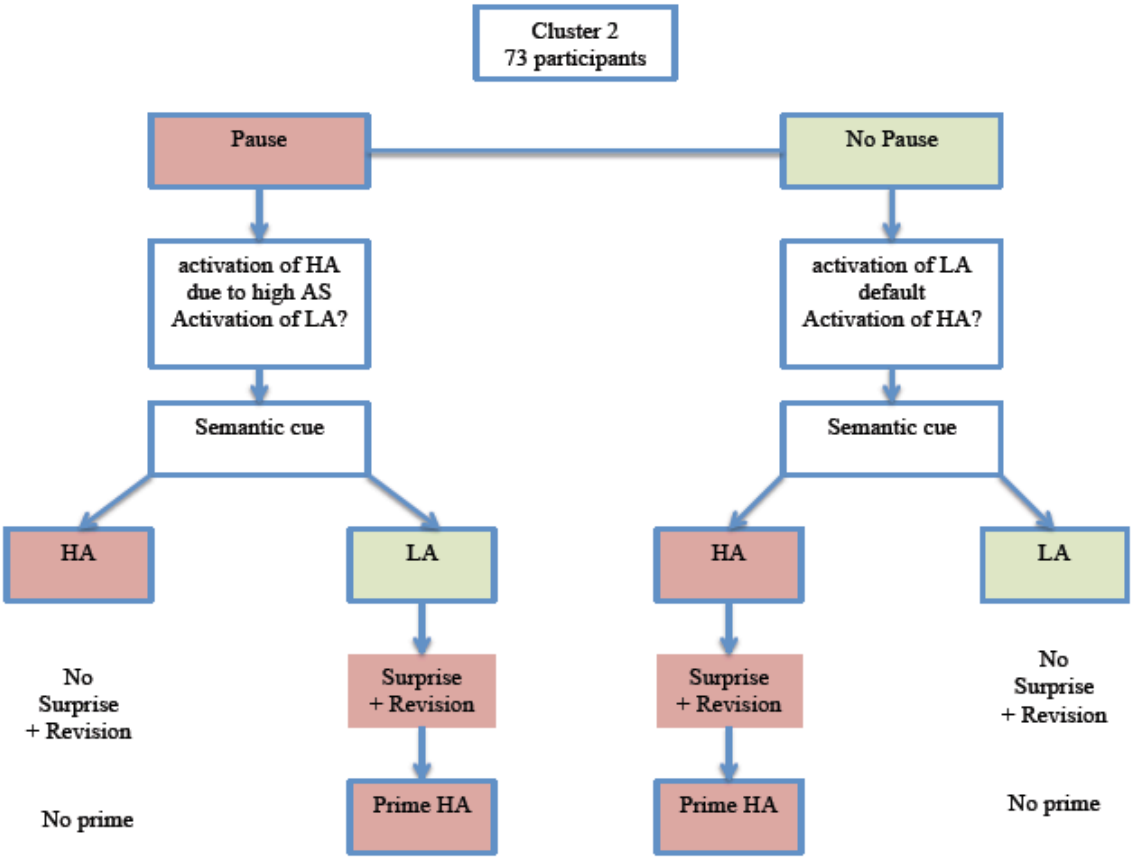
Now assume, no pause was present in the speech signal (1 b and d) which is in line with the low attachment preference. However, as mentioned above, this cluster might rely on prosodic cues less overall. This would impact on the weight assigned to certain cues and their impact on subsequent processing. Since participants in this cluster might rely on prosody less it could be possible that the alternative HA option was pursued alongside the default LA one. Subsequently, a low attachment semantic cue (1 d) is encountered which confirms one of the activated structures, namely syntactic LA. However, because prosody does not impact processing to a great extend, alternative structures

might have been activated, leading to competition. No surprisal value is assigned to the low attachment default option, therefore, no priming takes place.
Again, no pause was encountered in the speech signal followed by a surprising HA cue (1b). Based on the surprisal principle, revision should occur. The revision should result in a high attachment resolution of the RC. The type of revision necessary assumes *an early default cue such as a no-pause could be overruled by a later surprising cue such as semantic HA.* Thus, more HA target completions should have been observed. This was not the case. Again, one explanation of why priming did not occur would rest on participants' acoustic sensitivity and their processing strategies affected by this sensitivity. That is, since they rely less on prosody, they might have activated different syntactic alternatives. A surprising HA cue would add weight to the HA syntactic structure during revision. Since the disambiguating cue was encountered late, revision might take more time due to competition between alternatives. Thus, no priming was observed. On the other hand, the second explanation could be that revision occurred but was resolved towards low attachment and the default option does not prime. At this stage, I cannot rule out either option based on the data observed. Therefore*, an early default cue such as a no-pause could be overruled by a later surprising cue such as semantic HA **only if** the semantic HA cue adds enough weight in the time available for revision.*

2. Cluster 2 (73 subjects) prosody and plausibility interaction

The priming results for this cluster replicated the results found in the pilot study. Therefore, the principles put forward above should suffice to interpret these results. Figure 2 visualises the influence of the proposed principles during priming for cluster 2.

Figure 2



Thus again, assume a low attachment preference for RC attachment (1d) in English. So, low attachment should represent the default syntactic interpretation. Assume further the pause has been perceived by listeners in this cluster as salient and surprising due to higher acoustic sensitivity (AC). The pause therefore adds considerable weight to a possible HA attachment of the upcoming relative clause. The alternative LA structure might still be activated but competition is in favor of HA attachment based on the weight of the pause. Subsequent semantic information confirms the HA interpretation and no revision is necessary. Thus, priming was not observed. Then again, in (1c) a pause was perceived as surprising which added considerable weight to the HA syntactic structure. But subsequent semantic information confirms the LA structure (1c) and a conflict of cues is created. Revision needs to take place. The type of structural revision necessary states that in cases of cue conflict *an early surprising cue such as a pause could overrule a later default-supporting cue such* *as semantic LA.* Since the pause has been perceived as a strong cue, late semantic support for a different attachment might not be able to override the strong prosodic cue. Thus, the surprising prosodic HA cue primes subsequent HA target completions.

In case of prosodic low attachment (1 b and d), the early absence of a pause in the speech signal might not be perceived as surprising because it is in line with the default preference. It might no carry the same constraint weight as a pause so both syntactic alternatives (HA v. LA) might be activated albeit in slight favor of LA. Semantic information confirms low attachment (1d) which is in line with the default-supporting semantic cue. Thus, no revision is necessary and no priming due to surprisal was observed. On the other hand, no pause was present in the speech signal followed by semantic high attachment information. Since there was no strong prosodic evidence early on, a surprising plausibility cue overrules the weaker prosodic evidence resulting in high attachment priming triggered by surprising plausibility information. *Thus, an early default cue can be overruled by a later surprising cue.*

8.4. Conclusion

I have suggested some processing principles that might help explain the different priming results reported in this chapter. However, these principles are very speculative at this point in time because they rely on a number of assumptions. Much rests on the idea that acoustic sensitivity leads to different processing strategies. That is, if listeners do not rely on prosodic cues much due to lower acoustic sensitivity, they might rely more on simultaneous activation of alternative syntactic structures. Thus, they would be faced with more competition during processing. More competition during processing might lead to longer or more extensive revision. A research paradigm such as priming might not in all cases pick up the extent or indeed the resolution of the ongoing revision. This could be due to its implicit nature. Participants were not asked explicitly to resolve the ambiguity. Only after further research will it be possible to evaluate the true explanatory potential of the proposed processing principles.

After having investigated the interaction of prosody and plausibility using an off-line measure such as priming I will now turn to studying the time course of the interaction using an online method, namely EEG.

9. Event Related Potentials and relative clause attachment

9.1. Introduction

In Chapter (7) and (8) I have looked at relative clause attachment using the offline priming method. The previous priming experiments have shown that prosody and plausibility seem to interact in rather complex ways. That is, for some participants it was the cue that biased towards the dispreferred high attachment (*pause* or *semantic HA*) that primed subsequent target completion in case of cue conflict. For other participants, the picture was less clear. The only reliable priming result was found for the combined prosodic and plausibility high attachment condition (*pause + semantic HA*). The subject variable *acoustic sensitivity* suggested that the difference in priming results might be mediated by participants' acoustic discrimination ability. That is, higher acoustic sensitivity correlated with the type of interaction found. However, what the structural priming results cannot tell is how the attachment ambiguity is processed online. As detailed in Chapter 3.2, event related potentials (ERPs) were chosen as an online method because they offer some advantages.

A participants' EEG is recorded while many stimuli of the same kind are being presented. Recording the brain's electrical activity provides researches with a continuous measure of processing so any change in the signal is picked up immediately. Therefore, the temporal resolution of on-going processing is extremely good and makes this an ideal tool to study the time course of said processing. Much research has been done to study language processing using ERPs and some distinct components and effects have been identified which I will shortly summarise (for more detail see Chapter 2.4. of this thesis):

- Syntax-related components:
    - LAN (left anterior negativity) is an umbrella term for effects linked to processing of syntax. Found between 300-500ms after word onset LAN effects have mainly been observed for phrase structure

or morphosyntactic violations, e.g. *The ice cream was eat* (Friederici et al., 1996).

- P600 is a positivity that starts around 500ms after the onset of the word that marks a syntactic violation or an ungrammatical continuation (Hagoort, 2003). Positivities between 500 – 1200ms have been found with varying topographies. A P600 for outright violations such as a missing word in a sentence has been observed at posterior electrode sites. A positivity for complex or ambiguous sentences has also been found at (right-)anterior electrode sites. Thus, the topographical differences might be linked to differences between cognitive processes associated with revision and repair or complexity and ambiguity (Osterhout et al., 1994; Van Berkum et al., 1999; Kaan & Swaab, 2003).

- Plausibility-related component:
  - N400 is a negativity between 300 - 500ms after onset of the semantically anomalous word but has since been related to semantic context in a broader sense (Kutas & Federmeier, 2009).
- Prosody-related component:
  - CPS (closure positive shift) is a positivity that was found to reflect processing a prosodic boundary (Steinhauer et al., 1999).

Some previous ERP studies have looked at relative clause attachment. For the main results of the two most relevant studies see Table 1[12].

[12] Augurzky (2005) has investigated RC attachment in German. However, her results were very difficult to interpret and were therefore of limited predictive value. A short summary of her finding can be found in Appendix 4.

Table 1 Summary of ERP studies on RC attachment

Study	Language and task	Items and disambiguation	Effects
Kaan and Swaab (2003)	English Reading	**1. a. Preferred LA:** I cut the cake beside the pizzas that **were** brought by Jill. **b. Non-preferred HA**: I cut the cakes beside the pizza that **were** brought by Jill. **c. Ungrammatical:** I cut the cake beside the pizza that **were** brought by Jill. **d. Grammatical:** The man in the restaurant doesn't like the hamburgers that **are** on his plate. **e. Ungrammatical:** The man in the restaurant doesn't like the hamburger that **are** on his plate.	Posterior P600 for non-preferred HA and ungrammatical compared to preferred LA Anterior P600 for complex NP prep NP compared to single NP
Carreiras, Salillas and Barber (2004)	Spanish Reading	**2. a. Preferred HA:** El criado de la actriz que estaba **divorciado.** [The servant (masc) of the actress (fem) who was divorced (masc)]. **b. non-preferred LA:** El criado de la actriz que estaba **divorciada.** [The servant (masc) of the actress (fem) who was divorced (fem)].	Broad frontal P600 for dispreferred LA (500-700ms) Posterior P600 for dispreferred LA (700-1000ms)

The first was published by Kaan and Swaab (2003). They looked at RC attachment during reading in English in the context of the difference between revision, repair and complexity. Firstly, they investigated complex noun phrases (NP1 prep NP2) with an attached RC such as (1 a-c) where disambiguation is achieved by means of number agreement. They hypothesised that the non-preferred high attachment (1 b)[13] should induce revision processes compared to the preferred low attachment (1 a). This should be reflected by an anterior P600. In contrast, the ungrammatical condition (1 c) should trigger repair processes compared to (1 a) which should be reflected in a posterior P600. However, what they found were posterior P600 effects for both non-preferred and ungrammatical conditions compared to the preferred condition.

Secondly, they looked at structurally simpler single NP constructions such as (1 d and e) and more complex sentences (1 a-c).

They found that the more complex and ambiguous 2 NP conditions (1 a-c) elicited a more right anterior P600 compared to the simple NP conditions(2). Thus, based on their results, there seemed to be no difference between ungrammatical RC attachment and dispreferred HA attachment when the disambiguation was syntactic in nature (number agreement). Both elicit a posterior P600 compared to preferred LA attachment. No difference between revision and repair processes could be established. However, syntactic complexity and temporal ambiguity seemed to be taken into account during processing. This was reflected in an anterior P600 as a response to more complex ambiguous sentences compared to simpler unambiguous ones.

Carreiras, Salillas and Barber (2004) studied RC attachment in Spanish. In Spanish, the preferred attachment site of an ambiguous RC is NP1, thus Spanish shows a high attachment preference (Cuetos and Mitchell, 1988). In a reading experiment, they presented dispreferred low and preferred high attachment relative clauses that were always disambiguated syntactically (number and gender agreement between the last word and the attachment

[13] Recall that previous research has shown evidence for a low attachmet preference in English (Chapter 2.3.1. and 2.3.2.).

host). At the point of disambiguation they found a broadly and slightly frontal distributed positivity between 500-700ms and a more posterior positivity between 700-1000ms for dispreferred (LA) compared to preferred (HA) attachment.

The authors conclude that the earlier positivity should be compared to Kaan and Swaab's (2003) anterior positivity which is a response to complexity. The later posterior positivity in turn seemed to resemble the posterior, revision and repair related P600 of Kaan and Swaab's (2003).

To develop predictions for the current ERP study, Bögels et al. (2013) and Pauker et al. (2011) are of major interest. In both studies, experimental conditions were exploited that included a missing boundary and a misplaced one, which resembled to some extend the conditions under consideration here. That is, in the *pause + semantic LA* condition a pause is presented before the RC and later followed by semantic support for low attachment. In this case, the pause might be perceived as misplaced or superfluous. In the *no-pause + semantic HA* condition, no pause precedes the relative clause and this is followed by semantic support for high attachment. In this case, the listener might perceive a pause as "missing" from the speech signal. It was suggested by Pauker et al. (2010) that a misplaced boundary leads to more processing difficulty at the point of disambiguation as evidenced by a biphasic N400/ P600 pattern. A missing boundary on the other hand resulted in a smaller P600 which was interpreted as reflecting less extensive revision processes. In a follow up study, Bögels et al. (2013) reported clear P600 effects for both missing and misplaced boundaries. The authors concluded that misplaced as well as superfluous prosodic information leads to processing difficulty and revision is needed. The difficulty and extent of the revision seems more pronounced for a misplaced than for a missing boundary.

Based on the results cited above a number of predictions for the interaction of prosody and plausibility during RC attachment could be suggested.

When the prosodic high attachment cue becomes available (last syllable of NP2) I should expect to find the prosody-related CPS. This should be found for pause compared to no-pause conditions.

The second critical point in the present materials is the last word where the final plausibility cue to attachment becomes available and integration of all information needs to take place. There are several possible predictions for this point.

Kaan and Swaab (2003) as well as Carreiras et al. (2004) reported an effect of the language-specific attachment preference on RC attachment. Both found posterior P600 effects for non-preferred attachment (low in Spanish, Carreiras et al., 2004; high in English, Kaan & Swaab, 2003). Thus, I would expect to find posterior positivities in response to the dispreferred high attachment in English. However, Kaan and Swaab (2003) and Carreiras et al. (2004) solely utilised syntactic high or low attachment disambiguaition in their materials (see Table 1). In contrast, the materials used in the present experiment use prosody (pause v. no pause) and plausibility (semantic support for high vs. low attachment). Therefore, other effects that reflect the nature if the disambiguation are possible, e.g. earlier, plausibility-related effects.

Furthermore, the prediction of effects that are linked to the dispreferred nature of a high attachment cue could be supported on theoretical grounds by one of the previously proposed processing principles, i.e. the suprisal associated with a dispreferred attachment cue. Assuming English displays a low attachment preference and suprisal as a processing principle exists, either dispreferred HA cue should exhert influence during processing but this influence might be displayed in different ways depending on the individual cue.

The prosodic high attachment cue becomes availabe early on just before the relative clause. At this point the parser might pursue a HA structure based on the suprising and therefore weighty prosodic information. At the last word, sentence wrap-up is anticipated. Since high attachment is syntactically more demanding than low attachment it might be generally dispreferred in English and the parser could therefore experience processing or more specifically integration difficulty at the point.

On the other hand, a principle like suprisal should mark high attachment plausibility information as very prominent and weighty in terms of processing. Thus, whenever plausibility HA information is encountered the parser might also experience processing difficulty.

What kind of effect might be expected is somewhat difficult to predict since the cues are prosodic or plausibility-related in nature and not syntactic as reported before. It would be conceivable to expect a a late positivity (P600-like) in response to either dispreferred high attachment cue. However, an earlier negativity could also be envisaged since this effect has been linked to integration difficulty.

Importantly, I am looking at the interaction of prosody and plausibility. Of particular interest in this context were the two conditions where prosody and plausibility supported different attachments, i.e. *pause + semantic LA* v. *no-pause + semantic HA.* Some predictions might be derived from Bögels et al. (2013) and Pauker et al. (2011). That is, depending on the type of mismatch, a N400/P600 pattern for a misplaced boundary (*pause + semantic HA*) and a late positivity for a missing boundary (*no pause + semantic LA*) might be expected to emerge in the critical position of the last word where disambiguation via plausibility constraints takes place. However, as the priming results indicated, the perception of the boundary might depend on the participants' acoustic capabilities and consequently effect processing of the RC attachment.

The difference in priming effects during target completion seemed to be mediated by the participants' acoustic discrimination ability. It therefore seems reasonable to assume that this might play a role during online resolution of the RC attachment as well. Specific ERP components or effects were difficult to predict, as I could not find a precedent in the available literature. However, acoustic capabilities might effect the CPS component or the perceived conflict of cues further down the processing line.

To sum up I predict the following effects:

1. at the last syllable of N2: a CPS for pause compared to no pause conditions;
2. at the last word: early negativity and / or late positivity in response to the prosodic high attachment cue (pause) compared to the prosodic low attachment cue (no pause);

3. at the last word: early negativity and / or late positivity in response to the plausibility high attachment cue compared to the plausibility low attachment cue;
4. at the last word: late positivity for mismatching (*pause + semantic LA* v. *no-pause + semantic HA*) compared to matching (*pause + semantic HA* v. *no pause + semantic LA*) attachment cues.
5. the effects might be mediated by the between-subject variable *cluster*.

9.2. Materials

The same materials as in the priming Experiments 1 and 2 were used, that is, 120 items such as in 3, each with 4 conditions (see Chapter 5 for details):

(3)

a. Someone shot the servant of the actress [] who was very servile.

b. Someone shot the servant of the actress who was very servile.

c. Someone shot the servant of the actress [] who was very famous.

d. Someone shot the servant of the actress who was very famous.

Experimental items were randomised and divided into 4 initial presentation lists using a Latin-square approach. Thus, each lists contained 120 experimental items, 30 in each of the four conditions. A second randomisation created a further 4 lists which were used for a second EEG recording session (see participants section below). As a result, each participant was presented with the same item twice but in a different randomised order. 240 filler items were interspersed with the experimental items. There were always 2 filler items between experimental items. Each presentation list started with 3 filler items and each block started with 2 filler items. Each list was divided into 9 blocks of 40 soundfiles each (experimental items and filler items).

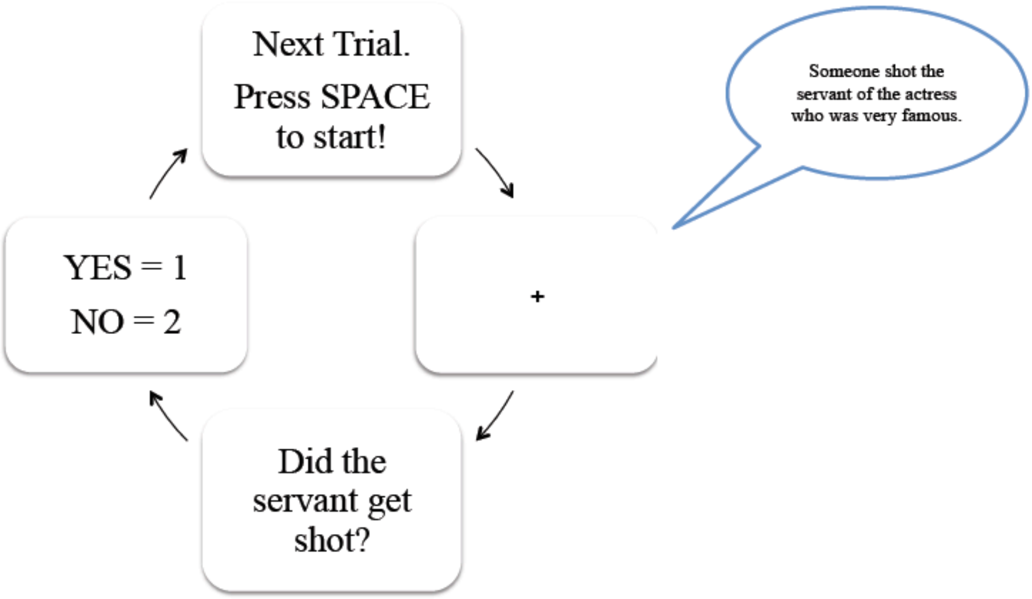
9.3. Design and procedure

Eprime software 2.0 was used to control stimulus presentation. The experiment had a fully crossed two-factor design with the factors *prosody* (pause vs. no pause) and *plausibility* (high vs. low attachment).

Participants were tested in a shielded, dimly lit room and heard the sentences over loud speakers. A written instruction sheet informed them about the experimental procedures. They were asked to listen to each sentence carefully for comprehension. All 120 experimental items and 80 of the filler items were followed by a simple comprehension task which related to the main clause (see Figure 1). No questions were asked relating to the relative clause in order to not draw attention to the experimental manipulation of the material. I also decided against a commonly used prosodic judgment task because such a task would draw attention to the critical factor prosody.

Figure 1 shows a typical experimental trial. The participant initiated each trial by pressing space. A fixation-cross appeared on the screen and the soundfile started between 400-700ms after trial onset to avoid expectancy related ERP responses. During the auditory presentation of the soundfile the participant looked at the fixation cross which remained on screen for 1000ms after the soundfile terminated to allow for time for the ERP effects to manifest at the last word. Each experimental trial and 80 filler trials were followed by a comprehension question. Participants replied to the comprehension question by pressing either 1 for yes or 2 for no on the keyboard in front of them. Subjects' responses to the comprehension question were recorded for later analysis. Between trials participants could take as much or little time as they needed to rest their eyes and get ready for the next trial.

Figure 1 Typical experimental trial



Before the experimental recording session started, participants performed the task in a practice block of 40 trials. A typical recording session lasted approximately 2 hours of which 1 hour was scheduled for electrode application and removal. Thus, it took participants typically 60 minutes to complete the experimental part.

9.4. Subjects

In the 2 priming experiments I could establish 2 groups of participants that showed opposing priming behavior. Therefore I decided to re-invite participants who had already participated in priming Experiment 1. Data collection for Experiment 1 finished in June 2011 and the EEG experiment data collection started in March 2012. Thus, there were at least 9 months between the two experiments.
An invitation to take part in the EEG experiment was send to all subjects who had taken part in priming Experiment 1. Out of all of the previous participants, 16 agreed to take part in the EEG experiment. Out of those 16

(8 female, mean age 26.1), 8 belonged to cluster 1 (2 female, mean age 28.9) and 8 belonged to cluster 2 (6 female, mean age 25.3). Each participant gave written consent and was paid £6 per hour.

I tested participants twice. This was done (a) to get an idea of test-re-test reliability of the ERP findings and (b) to evaluate potential learning effects resulting from repeated exposure to the same stimuli.

14 of the 16 participants returned for a second recording session with a mean of 38 days (range: 10-81 days) between recordings. For the second recording, participants were presented with the same items that were presented during the first recording but in a different presentation order.

9.5. EEG recording

EEG data were recorded using the 128-channel Biosemi Active Two EEG system (BioSemi, Amsterdam, Netherlands) and referenced to the common mode sense (CMS; active electrode) and grounded to a passive electrode.[14] I recorded from four additional electrodes – UltraFlat Active BioSemi electrodes – below and at the outer canthi of both eyes. The analogue signal was digitized at 512 Hz and filtered online at 100 Hz. Electrode offsets were kept between ±20 μV.

9.6. EEG data preprocessing

EEG data were pre-processed using Matlab 2011a and the open- source toolbox EEGLAB 11.0.2.1b (Delorme and Makeig, 2004). Firstly, data were re-referenced off-line to average reference and de-trended. Then, I removed noisy trials and channels without interpolation by visual inspection on a subject-by-subject basis (see Table 2). Consequently, between 105 and 120 trials per subject entered the data analysis.

[14] The layout of the electrodes on the head can be found in appendix 3.

Table 2 Descriptive analysis of removed noisy trials and channels

		EEG recording 1	EEG recording 2
noisy channels removed	Mean	11	13
	Std. Dev.	3	3
	Min.	5	1
	Max	16	20
Noisy trials removed	Mean	5	4
	Std. Dev.	4	3
	Min.	0	1
	Max	15	13

Secondly, the data were filtered before epoching. Two types of filters (high-pass causal filter and low-pass non-causal filter) were used because as Rousselet (2012) showed causal high-pass filtering seems to affect the onset of effects least and non-causal low-pass filtering has negligible effects on the time-course of ERP effects studied. In preparation for the independent component analysis I created 2 different datasets by applying 2 different filter settings. The first dataset was filtered using a 4th order Butterworth filters: high-pass causal filter at .16 Hz and low-pass non-causal filter at 30 Hz. The filter settings were based on filter setting reported previously (Pauker et al., 2010; Bögels et al., 2013). The second dataset was created using a high-pass causal 1 Hz and low-pass non-causal filter at 30 Hz. Subsequently, the data in both datasets was resampled at 500Hz and epoched between -500 before and 5750ms after sentence onset. Each epoch was then baseline corrected (-500-0ms).

Independent Component Analysis (ICA) was used for artifact correction, i.e. to identify and subsequently remove blinks. ICA relies on the assumption that the EEG signal recorded at any one electrode is made up of temporarily and topographically independent signals that are linearly summed up. ICA is capable of separating these independent signals based on their time course and topography (Jung et al., 2000). Thus, artifacts like blinks could be identified by their distinct time course and topography. ICA as implemented in the *runica* EEGLAB function (Delorme and Makeig, 2004; Delorme et al.,

2007) was computed on the 1Hz filtered data set. Using data sets with filter setting above 1Hz for ICA seems to produce clearer ICA components (Bieniek et al., 2013). Components corresponding to blink activity were identified by visual inspection of their scalp topographies, time-courses and activity spectra and their ICA weights were saved. ICA weights were then applied to the .16Hz causal Butterworth-filtered datasets (on a subject by subject basis) in order to ensure removal of the same components from both datasets. For subsequent ERP data analyses the .16Hz filtered dataset was used.

9.7. ERP analyses

EEG recording sessions were analysed separately. There were 2 reasons behind this decision. I wanted to test whether possible effects found during the first recording would replicate in a second recording. Also, 2 subjects from the first recording session did not take part in the second recording session. It was therefore decided to analyse the recordings separately to be able to keep the first recording of these 2 subjects.

Single subject averages of time windows (TW) of interest were computed for each condition. TWs of interest were identified based on previous studies (Bögels, 2011; Kaan & Swaab, 2003; Carreiras; 2004). For each of the TWs, the mean of the time window and a 200ms pre-stimulus interval were removed. The mean of the whole time window was removed to center the data. The pre-stimulus interval was removed to subtract stimulus-unrelated activity preceding the TW of interest.

One epoch of interest started at the onset of the last syllable of N2 before the relative clause assuming at this point the acoustic differences between pause and no pause conditions become available. 700ms epochs were extracted for each participant at the onset of the last syllable of N2 in each condition.

A second TW of interest was located at the onset of the last word of the sentence because at this point the semantic disambiguation cue became available. Firstly, 200ms epochs between 300-500ms after word onset were extracted for each participant in each condition. This epoch was thought to contain the previously described N400 time window. Additionally, 400ms epochs between 500-900ms after word onset were extracted for each participant in each condition. This epoch was expected to contain the previously described P600 TW. After plotting the data, a third time window of interest was identified, ranging from 100-300ms.

Statistical analyses were conducted at the group level using IBM SPSS Version 20.0.0. Repeated measure ANOVAs were run in the averaged TWs of interest.

At the onset of the last syllable of N2 before the point of interest, ANOVAs for lateral electrodes included the within-subject factors *prosody* (pause v. no-pause), *hemisphere* (left v. right) and 6 *regions of interest (RoI):* left anterior (electrodes 88, 89, 90, 91, 95, 96, 99, 100, 101, 102), left central (electrodes 105, 106, 107, 108, 109, 110, 114, 115, 116, 117, 118, 119), left posterior (electrodes 5, 6, 7, 8, 9, 16, 17, 18, 112, 113, 121, 122, 123, 124, 125, 126), right anterior (electrodes 67, 68, 69, 70, 73, 74, 75, 76, 77, 78), right central (electrodes 52, 53, 54, 55, 56, 57, 60, 61, 62, 63, 64) and right posterior (electrodes 29, 30, 31, 32, 34, 35, 36, 37, 38, 44, 45, 47, 48, 49, 50, 51). ANOVAs for midline electrodes included the factors *prosody* (pause v. no pause) and *RoI*: mid-anterior (electrodes 66, 84, 85, 86, 87, 98), mid-central (electrodes 1, 2, 33, 67, 97, 111), mid-posterior (electrodes 3, 4, 19, 20, 21). The factor *plausibility* (semantic support for high v. low attachment) was additionally included for the time windows following the onset of the last word, the second point of interest. The between subject factor *cluster* was included during the overall analysis.[15] In order to address violations of sphericity, the Greenhouse-Geisser correction was applied and corrected p-values will be reported. No multiple comparison correction, e.g. Bonferroni correction was applied. The first set of ERP analyses focused on the overall results and a second set concentrated on ANOVAs for each cluster separately.

[15] Significant interactions with the between subject variable can be found in appendix 4.

9.8. Results

9.8.1. Behavioral results

Firstly, the responses to the comprehension questions were analysed. Overall, the probability of making an error when responding to the comprehension question was low for both recordings (Mean Recording 1=.10, STD Recording1=.06; Mean Recording 2=.11, STD Recording 2=.07) demonstrating that participants were paying attention while listening to the sentences. Also, there were very slight descriptive differences between conditions detailed in Table 3.

Table 3 Probability of incorrect responses overall

	Recording 1		Recording 2	
Condition	Mean	Std. Dev.	Mean	Std. Dev.
Pause + semantic high attachment	0.08	0.05	0.08	0.05
Pause + semantic low attachment	0.11	0.09	0.13	0.09
No Pause + semantic high attachment	0.11	0.05	0.12	0.07
No Pause + semantic low attachment	0.11	0.06	0.12	0.05

As mentioned before, I found behavioral differences for the two clusters during the implicit priming experiments. Thus, I might expect to find behavioral differences in an explicit comprehension task as well. There were minor differences between the clusters in the two different recording sessions as detailed in Table 4 and Table 5.

Table 4 Probability of incorrect responses per cluster recording session 1

Recording 1	Cluster 1		Cluster 2	
Condition	Mean	Std. Dev.	Mean	Std. Dev.
Pause + semantic high attachment	0.06	0.03	0.10	0.05
Pause + semantic low attachment	0.11	0.07	0.11	0.10
No Pause + semantic high attachment	0.10	0.06	0.12	0.04
No Pause + semantic low attachment	0.10	0.06	0.13	0.07

Table 5 Probability of incorrect responses per cluster recording session

Recording 2	Cluster 1		Cluster 2	
Condition	Mean	Std. Dev.	Mean	Std. Dev.
Pause + semantic high attachment	0.08	0.04	0.08	0.06
Pause + semantic low attachment	0.12	0.11	0.15	0.07
No Pause + semantic high attachment	0.13	0.08	0.10	0.06
No Pause + semantic low attachment	0.13	0.05	0.10	0.06

However, the differences are purely descriptive in nature. Subsequent GEE analyses with a *binomial* distribution and *logit* link function were carried out modeling the error probability and using *plausibility*, *prosody* and *cluster* as factors in each of the recordings. No reliable main effects of or interactions between any of the factors were found (all ps > .07).

9.8.2. ERP results

I will report the results of the overall analysis first. The analysis will be reported separately for recording 1 which included 16 subjects and recording 2 which included 14 subjects.

9.8.2.1 ERP effects at the pause (onset of the last syllable of N2)

For the repeated measures ANOVA all items containing a pause and all items without a pause following N2 were averaged separately. In recording 1, repeated measure ANOVAs revealed no significant main effects or interactions of the factors (all p's > .07). The same was true for recording 2 (all p's > .33).

Thus, no significant ERP effects such as the closure positive shift (CPS) were found in response to the presence of a boundary in the speech signal. The implications of this will be considered in the discussion.

9.8.2.2. ERP effects at the point of semantic disambiguation (the last word)

9.8.2.2.1 Recording 1

In recording 1, repeated measure ANOVAs at lateral electrodes revealed a significant interaction of *prosody*region* in the earliest time wondow (F(2, 28) = 6.22; *Greenhouse-Geisser Epsilon* = .659; *adjusted* p=.016, Figure 2) as well as in the later time window (300-500ms) (F(2, 28) = 6.26; *Greenhouse-Geisser Epsilon* = .622; *adjusted* p=.015). The interaction indicated a negativity at posterior electrodes for pause compared to no pause conditions.

Also in the earliest time window and at midline electrodes, an interaction of *plausibility*region* was found (F(2, 28) = 4.56; *Greenhouse-Geisser Epsilon* = .593; *adjusted* p=.042, Figure 3), indicating a negativity at anterior electrode sites for high compared to low attachment.

No significant effects were found for the latest time window (*500-900ms*).

To illustrate the interactions, the ERPs time-locked to the onset of the last word (200ms baseline) for recording 1 are shown for pause compared to no pause conditions in Figure 2. In Figure 3 the ERPs time-locked to the onset of the last word for recording 1 are shown for high compared to low attachment (plausibility*region interaction). The ERPs were not filtered any further before plotting. Each figure shows 3 regions: anterior (F3, Fz, F4), central, (C3, Cz, C4) and posterior (P3, Pz, P4) as well as 2 hemispheres left (F3, C3, P3) v. right (F4, C4, P4) and midline electrodes (Fz, Cz, Pz).

Figure 2 Pause (red) v. no-Pause (blue), Recording 1

Grey areas highlight the time window of the significant simple effects of prosody

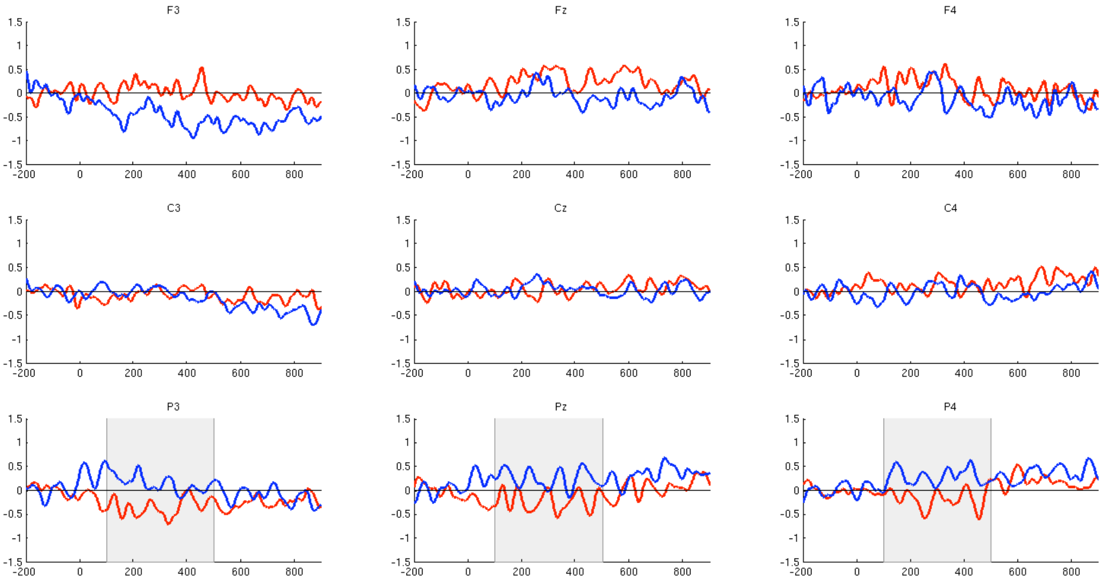
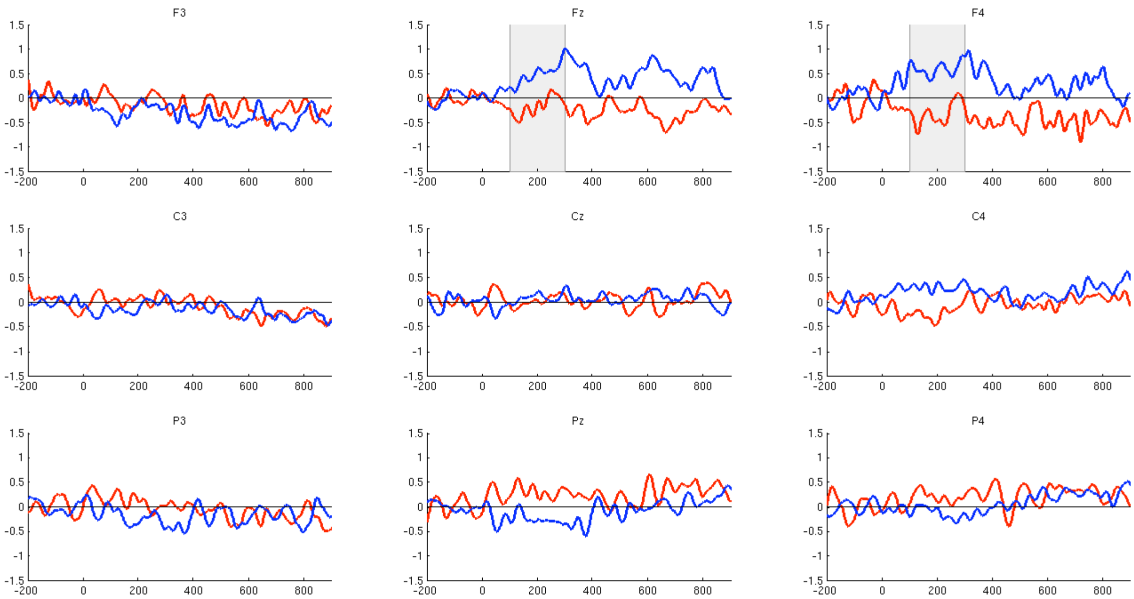


Figure 3 high (red) v. low attachment (blue), Recording 1

Grey areas highlight the time window of the significant simple effects of plausibility



To summarise, a posterior negativity was observed for pause compared to no pause conditions in two consecutive TWs. This might be interpreted as a sustained negativity between about 100ms to 500ms. This might be taken as an indication of sustained processing difficulty due to the presence of the pause. The prosodic cue for high attachment (*pause*) had been presented before the onset of the critical last word and should have cued the dispreferred HA structure. When the last word was encountered the previous presence of the pause seems to have induced processing difficulty independent of plausibility.
In the earliest time window, an anterior negativity was found for high compared to low attachment conditions. These results imply that the plausibility cue for high attachment (*semantic support for HA)* was perceived relatively fast and independently of other factors such as prosody and might have been recognized as diverging from the preferred low attachment. Consequently, both unexpected and possibly dispreferred cues to attachment might have influenced processing at the last disambiguating word independently and very early on.

9.8.2.2.2 Recording 2

In recording 2, no main effects of or interactions with the factor *prosody* at lateral or midline electrodes were found in the earliest time window.
An interaction of *plausibility*region* at midline electrodes were found ($F(2, 24) = 4.31$; *Greenhouse-Geisser Epsilon* = .587; *adjusted* p=.05, Figure 4) in the earliest time window. The interaction indicated a widespread negativity for high compared to low attachment with a maximum at anterior electrode sites.
In time window 300-500ms and 500-900ms, interactions of *prosody*plausibility* ($F(1,12)=11.94$, $p=.005$; $F(1,12)=6.57$, $p=.025$) and an interaction of *hemisphere*prosody*plausibility* ($F(1,12)=5.7$, $p=.034$) was found at lateral electrodes. The two-way interaction indicated firstly that in case of semantic high attachment the absence of a pause (*no-pause + semantic HA*) led to a positivity compared to the presence of a pause (*pause + semantic HA*).
In case of semantic low attachment (*pause + semantic LA* and *no-pause + semantic LA*) the presence of a pause led to a positivity compared to the absence of a

pause. In addition, the misplaced pause (*pause + semantic LA*) seemed to lead to a more pronounced positivity than a missing pause (*no-pause + semantic HA*). The three-way interaction indicated that this effect was more pronounced over the right than over the left hemisphere.

In Figure 4 the ERPs time-locked to the onset of the last word for recording 2 are shown for high compared to low attachment to illustrate the interaction of plausibility*region. Figure 5 shows the ERPs time-locked to the onset of the last word for recording 2 comparing high attachment in no pause and pause conditions to illustrate the interaction of prosody*plausibility. Figure 6 shows the ERPs time-locked to the onset of the last word for recording 2 comparing low attachment in pause and no pause conditions to illustrate the interaction of prosody*plausibility.

Figure 4 high (red) v. low attachment (blue), Recording 2

Grey areas highlight the time window of the significant simple effects of plausibility

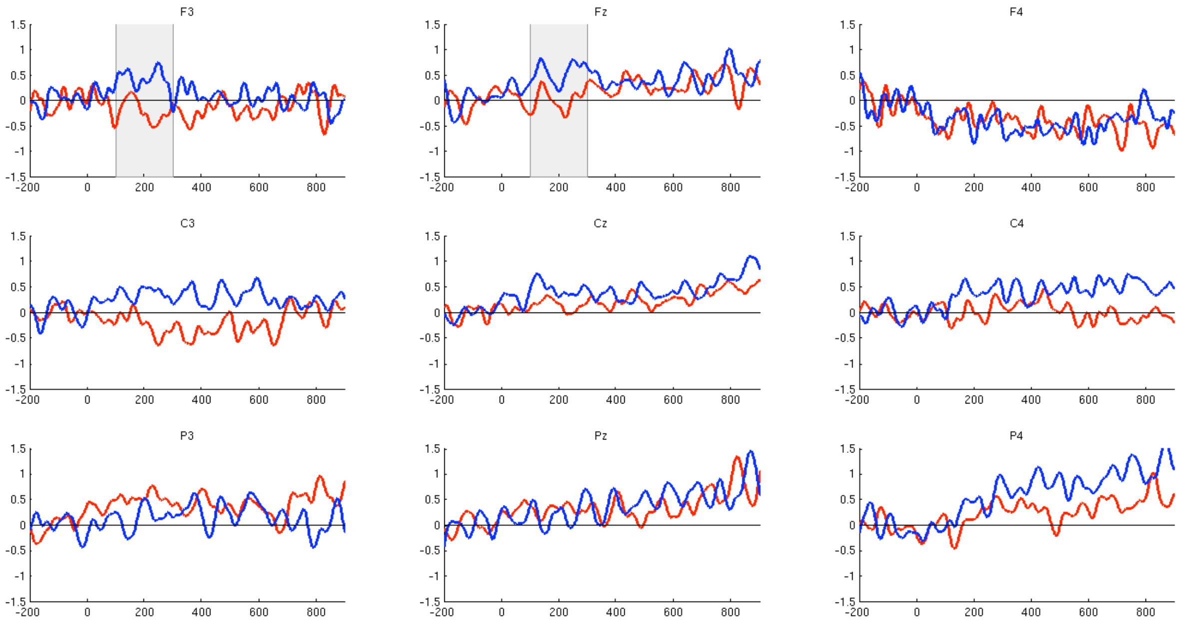


Figure 5 **Semantic high attachment** preceded by no-pause (red) v. pause (blue)[16], Grey areas highlight the time window of the significant interaction of prosody*plausibility

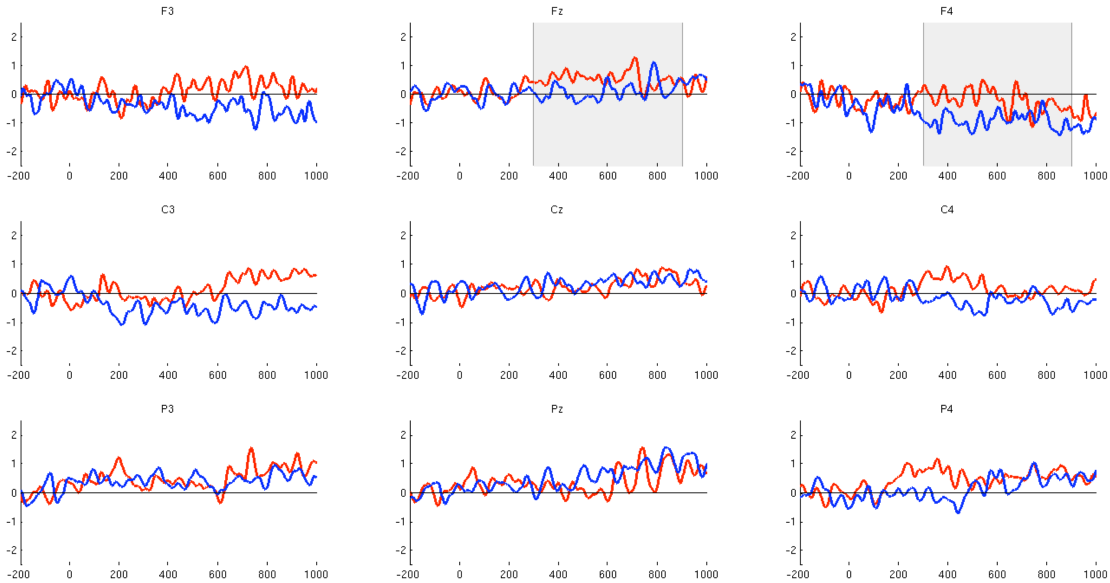
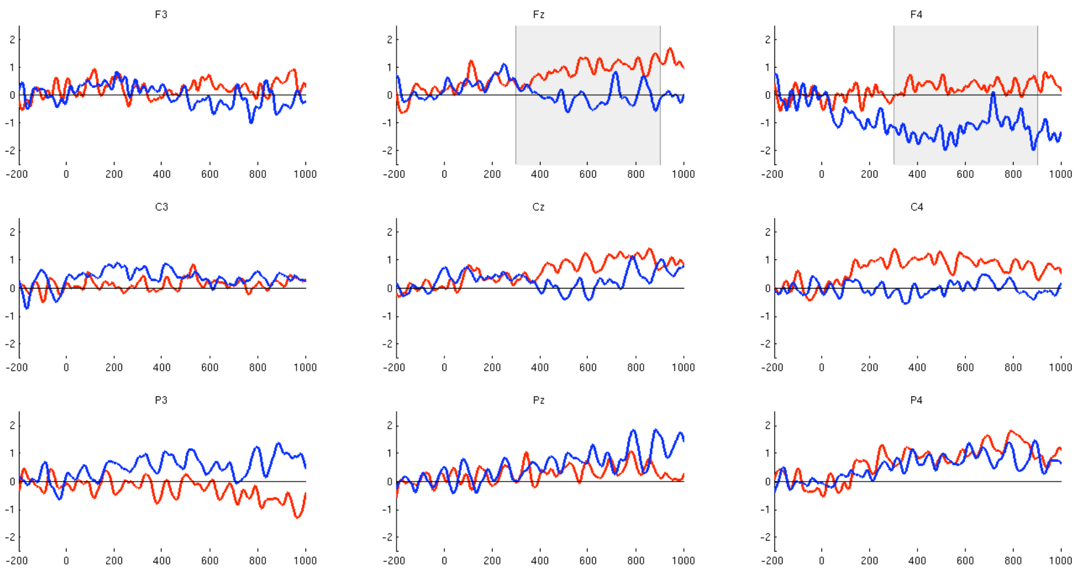


Figure 6 **Semantic low attachment** preceded by (red) v. no pause (blue)
Grey areas highlight the time window of the significant interaction of prosody*plausibility



[16] Red is always used to indicate conditions where cues support different attachment options and blue to indicated conditions where cues support the same attachment options

In sum, in recording 2 an early anterior negativity was found for high compared to low attachment plausibility information. A similar negativity had been found in recording 1. This result supports the previous assumption that the plausibility cue for high attachment (*semantic support for HA)* was perceived relatively fast and independently of other factors such as prosody. It might have been recognized as diverging from the preferred low attachment.

In contrast to recording 1, no effects of prosody were observed in this time window for recording 2.

However, an interaction of *prosody*plausibility* was found that started in the time window between 300-500ms and was still present in the later time window between 500-900ms. The positivity was more pronounced over the right than over the left hemisphere in both time windows. This could possibly imply that sustained processing difficulty was experienced.

9.8.3. Discussion of the overall ERP results for recordings 1 and 2

In this ERP experiment I investigated the online interaction of prosodic (pause v. no-pause) and plausibility (semantic support for HA v. LA) cues for relative clause attachment.

In the materials, a complex noun phrase (e.g. *servant of the actress*) was followed by a relative clause. The relative clause could either modify NP1 (high attachment to *the servant*) or NP2 (low attachment to *the actress*). The prosodic cues to attachment were a pause after NP2 which was assumed to support high attachment and no pause after NP2 which was assumed to support low attachment. The plausibility cues to attachment became available at the last word which disambiguated towards high or low attachment based on plausibility constraints, i.e. a servant is supposed to be more servile than an actress and an actress is supposed to be more famous than a servant. Both cues to attachment either agreed or disagreed in their in their support for high v. low attachment.

Some predictions were derived from previous studies into RC attachment.

Firstly, I expected to find the prosody related closure positive shift (CPS) at the boundary position at the second noun (Steinhauer et al., 1999; Bögels et al., 2011, 2013; Pauker et al., 2011).

Secondly, I hypothesised that the difference between preferred low attachment and dispreferred high attachment might be reflected in specific ERP effects.
Based on the type of disambiguation cue (prosody v. plausibility), I suggested to find different effects for the different types of attachment cues, i.e. early negativities and / or late positivities for dispreferred high attachment cues (pause v. plausibiity support for HA) compared to low attachment cues (no pause v. plausibility support for LA).
Thirdly, the interaction of prosody and plausibility was thought to be reflected in specific ERP effects. Late positivities were expected for mismatching prosodic and plausibility information (*pause + semantic LA* v. *no-pause + semantic HA*) compared to matching prosodic and plausibility information (*pause + semantic HA* v. *no pause + semantic LA*).
Before discussing the observed effects in detail, it is necessary to point out one important restriction associated with this ERP study and its results. When looking at the results and their interpretation it is important to keep in mind that the results are based on 16 participants in the first and 14 participants in the second recording. This was a low number of participants compared to Kaan and Swaab (2003) who used 26 and Carreiras (2004) who used 30 participants. Therefore, the results and their interpretation should be regarded as preliminary and should be taken with a pinch of salt until they can possibly be verified in future ERP experiments with a higher number of participants.
. To start with, it is worth mentioning that no CPS-like component was found at the boundary position at N2.This was unexpected, as the CPS has been reported repeatedly for processing of prosodic boundaries, even in different languages (Steinhauer et al., 1999; Pannekamp et al., 2005; Pauker et al., 2011; Bögels, 2011; Bögels et al., 2013). A possible reason for the absence of a CPS could be the subtle acoustic features of the stimulus materials as I have detailed in Chapter 5.2.1.
In the pause conditions, the speaker mainly relied on lengthening of the last syllable of the noun immediately preceding the relative clause to mark the boundary position. The pitch offset between pause and no-pause conditions at the last noun was statistically significant but rather small (Mean difference

6.9Hz). Additionally, the last noun before the RC was followed by a 250ms pause. These acoustic features taken together might not have sufficed to elicit a CPS.

Posterior negativity for Pause v. no Pause at the onset of the last word

In the overall analysis of the first recording a posterior negativity was found in the first (100-300ms) and second (300-500ms) time window for pause compared to no pause conditions. A similar posterior negativity was not found in the second recording.

The time course (approx. 100-450ms after word onset) and topography (posterior) might indicate that these two negativities in two time windows should actually be linked into one effect. The present posterior negativity might be linked to the notion of processing difficulties or increase processing demands due to the dispreferred nature of the HA structure which was built up following a suprising and prominent prosodic HA cue (pause).

Assuming low attachment is preferred in English, a pause has been encountered in the speech signal early on, which could be taken as a suprising cue to dispreferred high attachment. This pause might have made a high attachment of the upcoming RC more likely. High attachment might, nonetheless, still be dispreferred due to a language-specific low attachment preference. As the RC unfolds linguistic information is integrated into this context and reaches its final stages at the last word. Finally the last word is encountered. Possibly independent of its lexico-semantic content the parser might encounter processing difficulty or increased processing demands. This processing difficulty might be due to a high attachment structure that was built up but could still be dispreferred.

In contrast to recording 1, no effect of *prosody* was observed overall in the second recording. There are two possible explanations for this.

Remember that by the second recording, participants had already encountered the stimulus material twice, the first time during the priming experiment and the second time during the first recording session. The repeated exposure, although weeks apart, might have given participants opportunity to accommodate to the presence of a pause. Thus, based on previous exposure, participants might not anticipate processing difficulties based on the pause cue alone.

Alternatively, the ow number of participants (14) in the second recording did not suffice to elicit any prosody-related effects.

Anterior negativity for HA v. LA

At the last word another negativity was found. This negativity was reliably elicited by high compared to low attachment at anterior electrode sites in both recordings in the earliest time window (100-300ms). The negativity was observed for high compared to low attachment conditions.

Low attachment seems to be the preferred attachment option in English. Encountering a high attachment cue might therefore be dispreferred and lead to processing difficulty. The early time window might suggest rapid access to the kinds of semantic representations associated with the high attachment plausibility cue.

At face value it seems tempting to relate this finding to already documented ERP effects such as the early left anterior negativity (ELAN) or its later instantiation, the LAN. To interpret the present negativity as an effect similar to an (E)LAN seems problematic at closer inspection. Indeed, E/LAN effects have been observed and documented for violations of morpho-syntax or the phrase structure of a sentence (Friederici et al., 1996). In the present experiment, there was no outright syntactic violation present. Rather, a difference in attachment between high or low was present. Thus, I might be looking at a difference between a preferred and a non-preferred grammatical structure. Additionally, no (early) negativity has been reported in connection with relative clause attachment as yet.

To conclude, this negativity might reflect processing difficulties associated with the suprising and dispreferred high attachment cue. No more in depth interpretation will be attempted at this point, as the evidence is not deemed sufficient enough to support further conclusions.

Anterior positivity LA (Pause v. no Pause) v. HA (no Pause v. Pause)

An anterior positivity was found during the second recording in two consecutive time windows (300-500ms and 500-900ms). The positivity was elicited by conditions supporting different attachment options compared to the conditions supporting the same attachment i.e. *pause + semantic LA* compared *no pause + semantic LA* and *no-pause + semantic HA* compared to *pause + semantic HA.*

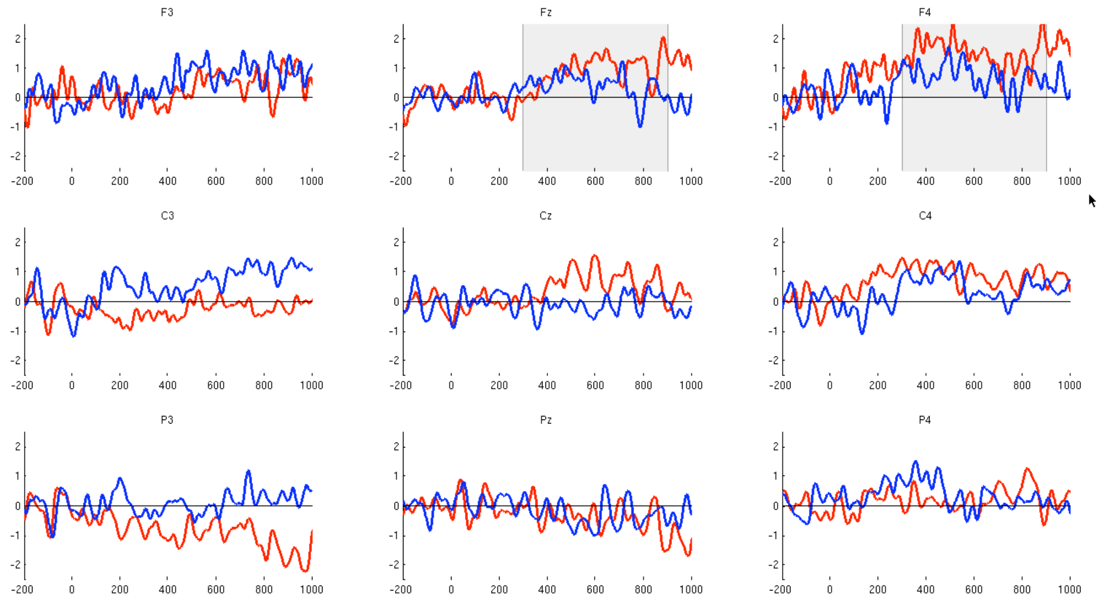
It is difficult to interpret the positivity found here as a classic P600 effect. P600s were mostly reported at centro-posterior electrode sites and they were commonly linked to syntactic violations. In previous research on RC attachment, syntactic disambiguation and violations were used (Kaan & Swaab, 2003) and a posterior P600 was reported. The present materials did not involve an out-right syntactic violation. Rather, prosody was used to cue a possible syntactic analysis early on and plausibility was used to disambiguate the intended RC attachment at the final word. So a match or mismatch of cues to RC attachment was employed. Thus, it is not entirely surprising that no posterior P600 was found.

Anterior positivities in contrast have been reported for RC attachment, mainly in response to complexity and ambiguity (Kaan Swaab, 2003). That is, a anterior P600 was reported for single NP RC structures compared to complex NP prep NP RC structures. As the present material only included complex nun phrases it seems a little premature to attempt a more in-depth interpretation along the same line as Kaan and Swaab (2003).

However, the disambiguation in the present materials was semantic in nature. The late positivity effect might therefore reflect incorporating plausibility information that relates to the complex noun phrase into a structure that was prosodically cued. Thus, revison processes might be going on. A similar observation has been reported by Pauker et al. (2011) and Bögels et al. (2013) who found P600 effects for misplaced and missing prosodic information compared to correct prosodic information. However, Pauker et al. (2011) suggested that there might be a difference between a misplaced boundary and a missing boundary. They suggested that the type of revision (deleting a misplaced pause v. inserting a missing pause) might lead to different P600 effects.

In the current study, the observed positivity seemed somewhat more pronounced for the misplaced pause (*pause + semantic LA*) compared to the missing pause (*no pause + semantic HA*) (see Figure 7). However, the differences turned out to be not significant in both time windows (allp's>.102).

Figure 7 Effect of a misplaced pause in the low attachment condition (red) v. effect of a missing pause in the high attachment condition (blue). Differences between *pause + LA and no-pause + LA v. no-pause + HA and no-pause + HA* were calculated and are shown in μV.



One last question needs to be addressed. That is, why was the anterior positivity found in the second but not in the first recording? Again, as mentioned above, by the second recording session, participants have encountered the materials twice before. Although many weeks apart, the presentation of matching prosodic and plausibility cues might have strengthened their possibly preexisting positive association. That is, a representation of pause and semantic support for high attachment might have already existed but might have been weak because of its dispreferred status. Since participants were repeatedly exposed to this combination, a pause and semantic support for high attachment might have become more strongly associated with an overall high attachment representation of the relative clause.

Conversely, no pause and semantic support for low attachment might have already been associated with an overall low attachment representation of the relative clause. This attachment might even be preferred. Thus repeated exposure would strengthen the association even further.

During the previous experiments (priming and recording 1), mismatching prosodic and semantic cues were presented alongside matching ones which could have led to a negative association being emphasized. That is, the positive evidence of the match might have highlighted the mismatch. Therefore, I might be looking at some type of learning through repeated exposure.

However, some caustion is advised. The fact that no indication of cue conflict was found during the first recording, should not be taken as evidence that no attempt was undertaken to resolve the conflict. It is entirely possible that the resolution started in the first recording but due to the low number of participants the analysis did not reach significance. Therefore the absence of the effect in the analysis is no evidence for the absence of the effect. This argument does not falsify the exposure-based learning idea.

One of the interesting results, reported for the priming experiments was the detection of two clusters of participants. The clusters displayed opposing priming results and this might be linked to their acoustic discrimination ability. Participants were invited to take part in the ERP study based on their previous priming results. It seems therefore reasonable to analyse the data separately for each cluster to investigate whether some differences in the ERPs might be present.

9.8.4 ERP effects for Clusters 1 and 2

9.8.4.1 ERP effects for cluster 1, Recording 1 and 2

The (supposedly) acoustically less sensitive cluster 1 did not show any effects in any time window during recording 1 ($p>.071$).

In contrast, during the second recording, this cluster revealed a broad negativity for high compared to low attachment in the earliest time window ($F(1,7)=11.89$; *adjusted* $p=.011$) and a positivity in the latest time window for conditions supporting different attachment options (*pause + semantic LA* v. *no-pause + semantic HA*) compared to the conditions supporting the same attachment (*pause + semantic HA* v. *no-pause + semantic LA*) over anterior lateral electrode sites ($F(1)$=4.23; *adjusted* p=.05).

9.8.4.2 ERP effects for cluster 2, Recording 1 and 2

During the first recording, the (supposedly) acoustically more sensitive cluster 2 revealed a negativity for the pause compared to no-pause conditions at lateral posterior electrodes in th earliest time window (*F(2,14)=7.65; Greenhouse-Geisser Epsilon* = .736; *adjusted p=.013*) and the time window ranging from 300-500ms (*F(2,14)=4.95; Greenhouse-Geisser Epsilon* = .612; *adjusted p=.049*). The second cluster also replicated the interaction of *plausibility* and *hemisphere* in the earliest time window (*F(1,7)=14.93; Greenhouse-Geisser Epsilon* = 1; *adjusted p=.006*). That is, a negativity was found for semantically supported high compared to low attachment that was more pronounced over the right than over the left hemisphere.

In the second recording, this cluster showed a negativity for semantic high compared to semantic low attachment. This negativity reached significance over the posterior region of the right hemisphere during the earliest time window (*F(2,10)=5.72; Greenhouse-Geisser Epsilon* = .621; *adjusted p=.048)*. This cluster also revealed a positivity at anterior lateral electrodes ($F(1)$=7.30, *adjusted* p=.043) with a maximum at anterior midline electrode sites ($F(1)$=8.63, *adjusted* p=.030) in the second time window and the latest time window ($F(1)$=6.42, *adjusted* p=.052).

9.8.4.3 Discussion

In two priming experiments (Chapters 6 and 7) I have identified clusters of participants that showed different behavioral responses towards the prosodic and plausibility cues to RC attachment. It was therefore decided to re-invite participants based on their priming results.

Interpreting the results of the ERP cluster analysis relies on 8 participants for both recordings in the (supposedly) acoustically less sensitive cluster 1.

8 participants for the (supposedly) acoustically more sensitive cluster 2 took part in in recording 1 and 6 participants in recording 2. Again, this is a very low number of participants.

However, the most interesting result of the ERP analysis is that differences between the clusters were observed. In sum, the (supposedly) acoustically more sensitive cluster replicated the results observed in the overall analysis whereas the (supposedly) acoustically less sensitive cluster did not. The results suggest that acoustic sensitivity might play a role in how participants process prosodic and plausibility cues to RC attachment.

Posterior negativity for Pause v. no Pause

In the overall analysis of the first recording posterior negativities were found in the first (100-300ms) and second (300-500ms) time window for pause compared to no pause conditions. It was suggested to interpret these as a linked effect.

During the analysis of the different clusters it was found that the effect was significant for the (supposedly) acoustically more sensitive cluster but not the (supposedly) acoustically less sensitive cluster. This implies that the overall effect might have been driven by the acoustically more sensitive cluster.

I argued that the early negativity might be linked to the notion of processing difficulties or increase processing demands due to the dispreferred nature of the HA structure which was built up following a suprising and prominent prosodic HA cue (pause). That is, a high attachment structure was built up based on the dispreferred pause cue. Since the second cluster seems to be acoustically more sensitive it might be possible that those participants experienced difficulties because they have perceived the cue as prominent and used it to built up a syntactic structure. No posterior negativity for pause compared to no-pause conditions was found for either cluster in the second recording.

I have argued above that the repeated exposure to the prosodic cue might have given participants opportunity to accommodate to the presence of a pause. Thus, based on previous exposure, acoustically more sensitive participants might not predict processing difficulties based on the pause cue alone.

Conversely, the acoustically less sensitive cluster might not rely on prosodic cues as much and might therefore not experience (or not reveal) processing difficulties due to the pause cue. Then again, the by-cluster analysis decreased the number of participants, thus decreasing the statistical power of the analyses.

Anterior negativity for HA v. LA

Overall, a negativity was reliably elicited by high compared to low attachment semantic cues in both recordings in the earliest time window (100-300ms). The negativity was observed for high compared to low attachment conditions. This negativity was also observed in the acoustically more sensitive cluster 2 in recording 1 but not in the acoustically less sensitive cluster 1. Although I have indicated that this negativity might reflect processing difficulties associated with the unexpected and dispreferred high attachment cue, some caution is advised. That is, in recording 2, the acoustically more sensitive cluster 2 revealed a posterior negativity whereas the acoustically less sensitive cluster 1 showed a broadly distributed negativity for high compared to low attachment conditions. The differences in the distribution of the effect are quite puzzling and as yet unexplained. Therefore, a more thorough interpretation at this point might be ill-advised until more research is carried out.

Anterior positivity mismatching compared to matching cues

Overall, an anterior positivity was found during the second recording in two consecutive time windows (300-500ms and 500-900ms). The positivity was elicited by conditions supporting different attachment options compared to the conditions supporting the same attachment i.e. *pause + semantic LA* compared *no pause + semantic LA* and *no-pause + semantic HA* compared to *pause + semantic HA.* I have suggested that present anterior positivity might be interpreted as a sustaineds processing difficulties due to revision processes. The relative clause modifies the complex noun phrase and the late positivity might indicate the resolution of conflicting prosodic and semantic information about which of the nouns is being modified. The presence of the interaction in recording 2 but not in recording 1 was interpreted as some type of exposure based learning. It was argued that the positive association of matching cues (*pause + semantic HA* and *no-pause + semantic LA*) might have been strengthened and the negative association between mismatching cues (*pause + semantic LA* and *no-pause + semantic HA*) possibly weakened.

Intriguingly, the clusters displayed differences in terms of the timing of the effect. The interaction of *prosody8plausibility* was significant in both time window (300-500ms and 500-900ms) for cluster 2. In contrast, the interaction reached significance only in the latest time window (500-900ms) for cluster 1.

From this, one might conclude that both clusters seem to display some exposure based learning. However, the acoustically more sensitive cluster 2 seems to display processing difficulties due to mismatching information earlier (300-500ms) and the difficulties might be more sustained. The acoustically less sensitive cluster shows some processing difficulties in the latest time window (500-900ms) only. Thus, the acoustic ability of participants might influence the timing and severity of possible processing difficulties due to conflicting prosodic and plausibility information.

9.9. Conclusion

To conclude, the present data suggest a number of inferences.

Firstly, more research might be necessary to clarify to conditions under which a CPS can be elicited. For the current materials it was not found.

Secondly, at the point of disambiguation different processes seem to take place. These processes might be linked to context-dependent prediction (in the case of pause v. no pause conditions), language-typical attachment preferences (in the case of low v. high attachment), and some type of exposure-based learning (in the case of matching and mismatching cues to attachment).

Thirdly, the effects might be mediated by participant specific capabilities such as their acoustic discrimination ability.

These results are both exciting and encouraging. However, before they might be replicated using a larger number of participants, some caution as to the validity of the interpretation is advised.

10. General discussion

In this thesis I have been investigating the interaction of overt prosodic (*pause to indicate high attachment* v. *no-pause to indicate low attachment*) and plausibility cues (*semantic support for high attachment* v. *low attachment*) to relative clause attachment in English. The overall aim of this thesis was to establish what type of interaction could be expected when prosodic and semantic cues support the same or different attachment options in a RC attachment ambiguity as shown in 1 (a-d):

1 a. *Someone shot the servant of the actress* [Pause] *who was very servile.*

b. *Someone shot the servant of the actress who was very servile.*

c. *Someone shot the servant of the actress* [Pause] *who was very famous.*

d. *Someone shot the servant of the actress who was very famous.*

To address the interaction, off-line as well as on-line methods were employed in order to assess the relative and combined contribution of the individual cues in a more thorough way.

Priming relies on repetition. A prime sentence is presented and the resolution of the constraints inherent in the prime should be reflected when a similar structure is presented in the subsequent target sentence. Therefore, priming could give an indication of the interaction of the constraints during the presentation of the prime.

Priming was also used because of its implicit and unbiased nature, as especially prosodic manipulations of speech materials might be susceptible to task effects when more explicit tasks such as comprehension questions were utilised (Lee & Watson, 2011).

Event-related potentials (ERPs) were employed as this online electrophysiological measure enables researches to study language processing as it unfolds over time. It will therefore give an indication of the interaction of the constraints while they are being processed.

The off-line method of priming was exploited in three experiments.

The first experiment was thought to be an exploratory pilot study. In the pilot study, prosodic and plausibility cues were found to interact in a surprising way. That is, without a pause before the prime-RC (prosodic support for low-attachment), plausibility cues (semantic support for either high or low attachment) showed reliable priming effects in the expected direction: more HA target completions after semantically HA biased primes and more LA target completions after semantically LA biased primes.

However, when the prime-RC was preceded by a pause (prosodic support for high-attachment), the effect of plausibility was *supressed*. I found more high attachment target completions after low attachment plausibility primes.

The pilot study was followed by the first main priming experiment. This priming experiment used a subtler prosodic manipulation in the materials, a larger number of stimuli (120 sets of items) and participants (72). Since there were acoustic differences between the materials used in the pilot study and those used in the subsequent priming experiments, subject specific variables were recorded for each participant, i.e. their acoustic discrimination ability and their listening memory span.

The results coming from the first major priming experiment turned out to be complex. Firstly, the overall analysis returned no significant main effects or interactions. Instead, two groups of participants were discovered in a cluster analysis. These groups showed opposing trends in the priming data. Specifically in the pause conditions, the clusters responded to the match (*pause + semantic HA*) or mismatch (*pause + semantic LA*) of the prosodic and semantic cues differently. Cluster 1 showed more HA target completions after matching cues (*pause + semantic HA*). Cluster 2 revealed more HA target completions after mismatching cues (*pause + semantic LA*). Cluster 2 therefore replicated the result found in the pilot study. No priming effects were found for either cluster in the no-pause conditions in this experiment.

There was a descriptive difference in participants' acoustic discrimination ability, cluster 2 showed a higher acoustic sensitivity than cluster 1. However, the difference did not reach statistical significance. No difference was observed for the variable listening memory span.

It was pointed out that a cluster analysis is designed to find yet undiscovered patterns in the data. However, whether the pattern that was found represents a true effect can only be determined by exactly replicating the experiment with a new set of participants. This was done in main priming Experiment 2.

In the second main priming experiment (72 subjects), two clusters of participants were found via a cluster analysis.

These new clusters showed the same priming effects in the pause conditions, i.e. cluster 1 revealed more HA target completions after matching cues (*pause + semantic HA*) and cluster 2 had more HA target completions after mismatching cues (*pause + semantic LA*). Both clusters differed in their priming results in the no-pause conditions. When comparing the two no-pause conditions, cluster 1 showed an unexpected effect, i.e. more high attachments for matching low attachment cues (*no-pause + semantic LA*). In cluster 2, in comparison, the mismatch (*no-pause + semantic HA*) primed strongest. In sum, cluster 2 replicated the result found in the pilot study.

Again, a descriptive difference in terms of acoustic discrimination ability was found. The difference was marginally significant (p=.07). Again, no difference was observed for the variable listening memory span.

Since very similar clusters were found in two identical priming experiments, it was decided to combine the datasets to increase the statistical power of the analysis.

In the now combined dataset, the analysis showed that the acoustically more sensitive cluster 2 replicated the result found in the pilot study. In case of cue conflict (*pause + semantic LA* and *no-pause + semantic HA*), it seems to be the surprising and dispreferred high attachment cue (*pause* or *semantic HA*) that influences subsequent target completion.

The acoustically less sensitive cluster 1 revealed that the priming effect in the pause conditions was driven by combined support for high attachment (*pause + semantic HA).* In contrast, the result in the no-pause conditions showed no consistent priming effects.

At this point, it can be concluded that the priming experiments have given some indication about the constraints that are strongly activated during prime processing to influence subsequent target completions.

To get a clearer picture about the actual time course of the interaction of prosody and plausibility, an ERP study was run. Importantly, the same materials that were used as prime sentences in the main priming Experiments 1 and 2 served as experimental items in the ERP study.

Participants that had already taken part in the first main priming experiment were re-invited to participate in the ERP study. Since two different clusters with different acoustic discrimination abilities and opposing priming results were discovered, this procedure seemed promising. I expected to find differences in the ERPs based on the differences in the priming results. Importantly, subjects were recorded twice to investigate test-retest reliability and to look at whether repeated exposure to the same materials led to some kind of learning effect.

There were two critical points in the materials. The first point started at the onset of the last syllable of N2. Here the acoustic differences became available. The second critical point started at the onset of the last word of the structure. The semantic disambiguation cue became available here.

At N2 position a time window of 700ms was extracted for each participant and analyses were carried out on the group level. No differences between pause and no-pause conditions were found in either recording. The prosody-related CPS component was not observed for the present materials.

Three time windows of interest were identified at the last word, i.e. 100-300ms, 300-500ms and 500-900ms. Overall inferential analyses revealed differences between the two recordings in these time windows.

In the first recording, a posterior negativity was found in the first and second time window for pause compared to no-pause conditions. It was suggested to interpret this negativity as an indication of sustained processing difficulty due to the presence of the pause. That is, processing difficulty might be encountered because a high attachment structure was built up based on the pause cue. This structure could still be dispreferred. Thus when the last word is encountered, the early negativity might reflect integration difficulty because of the dispreferred nature of the current structure.

A similar negativity was not found in the second recording. The absence of this effect might be an indication of exposure-based learning taking place. The repeated exposure to the same materials might have given participants opportunity to accommodate to the presence of a pause. Thus, based on previous exposure, participants might not anticipate processing difficulty based on the pause cue alone.

During the analysis of the individual clusters the interaction of *Prosody*Region* was found in the two earlier time windows for the acoustically more sensitive cluster only. Therefore, it seems that the effect of prosody observed in the overall analysis might have been driven by the acoustically more sensitive cluster.

An anterior negativity was found in the earliest time window (100-300ms) in both recordings for high compared to low attachment conditions. The interpretation of this negativity as an (E)LAN was rather difficult since no outright syntactic violation was present in the current materials. Rather, a difference is described in terms of high or low attachment. Specifically, the materials included a difference between a preferred and a non-preferred grammatical structure. This negativity might reflect processing difficulty associated with the unexpected and dispreferred high attachment cue.

During the analysis of the individual clusters an early effect of plausibility was found for the acoustically more sensitive cluster only. Thus, this cluster might also have driven the effect of plausibility observed in the overall analysis of recording 1. The results for recording 2 for this negativity revealed some differences in topography between the clusters. Based on the limitations associated with this experiment I refrained from interpreting this difference until more experimental evidence might become available in the future.

In the overall analysis, an anterior positivity was found during the second recording in two consecutive time windows (300-500ms and 500-900ms). The positivity was elicited by conditions supporting different attachment options compared to the conditions supporting the same attachment. i.e. *pause + semantic LA* compared *no pause + semantic LA* and *no-pause + semantic HA* compared to *pause + semantic HA.* It seems that listeners experience processing difficulty at the last word when a prosodically supported structural interpretation clashes with subsequent semantic information.

This positivity could be interpreted as a complexity and ambiguity related P600 effect. As the disambiguation in the present materials was semantic in nature, the P600 effect might reflect some sort of processing difficulty. A similar positivity was not found in the first recording. Again, this difference might reflect some type of learning through repeated exposure.

Both clusters showed the anterior positivity in the conditions where cues to RC attachment favoured different attachment options compared to when they supported the same attachment option. However, the analysis revealed differences in timing of the effect. The acoustically more sensitive cluster showed this positivity already in the earlier time window (300-500ms). In the acoustically less sensitive cluster, the positivity was found only in the late time window (500-900ms). I suggested a cluster-dependent processing of prosodic and semantic cues to RC attachment, which might be linked to listeners' acoustic capabilities.

Taken together, the results of the priming and ERP experiments suggest a number of conclusions.

The first important finding of this thesis is that prosodic and semantic cues interact in complex ways. That is, when cues support different attachment options, the dispreferred high attachment cue (*pause* or *semantic* HA) in the prime leads to more high attachment target completions. Thus, the results revealed by the offline method priming, indicate that the dispreferred cues seem to be highly activated during priming. The activations seemed to have been strong enough to increase a dispreferred target completion.

When using an online measure, the cues to the dispreferred structure induce processing difficulty independent of each other. That is, at the point of disambiguation (the last word) processing difficulty is in the first instance observed for dispreferred cues in general as suggested by early negativities for high compared to low attachment and pause compared to no-pause conditions. It seems that the preference for a certain structure, i.e. low attachment in English, is reflected in early ERP components. This could be taken as evidence that a preference for RC attachment exists (LA) and that cues to indicate dispreferred attachment (*pause* and *semantic HA*) are recognised and assessed very quickly. Importantly, no indication of revision as observed in P600 effects was found in the first recording. It seems that a certain amount of exposure is necessary for the parser to learn and differentiate between matching information that eases efficient processing and mismatching information that hinders efficient processing.

The second important finding of this thesis is that the interaction depends on the prominence of the acoustic high attachment cue or the acoustic sensitivity of the participants respectively.

That is, when the acoustic cue is very prominent such as in the pilot study, the priming effects occur in all participants. When the acoustic cue is rather subtle, the priming effects occur only for participants with higher acoustic sensitivity. This is supported by the results of the ERP study. ERP effects were found in the overall analysis, but they seemed to be driven by the acoustically more sensitive participants who seem to pick up subtle prosodic cues and rely on them during processing. This is evidenced by the fact that all interactions observed in the first and most observed in the second recording reached significant in this cluster only.

In contrast, the priming results and the ERPs found for the acoustically less sensitive cluster offer limited opportunity to draw conclusions about the interaction of prosody and plausibility since (1) priming effects were only found for matching high attachment cues and (2) ERP effects were observed in the second recording but especially the early negativity for high compared to low attachment was difficult to interpret.

I will subsequently concentrate on the results of the pilot study and the results displayed by the acoustically more sensitive cluster. Those results offer the starting point for discussing the interpretation of the observed interaction between prosody and plausibility.

Based on the pilot study, I suggested two principles that might account for the observed interaction. The principles involved the *surprisal associated with the high attachment cue (pause v. semantic support for high attachment)* and *the type of structural revision necessary*.

The results of the two main priming experiments and their combined analyses suggested (1) the existence of a third processing principle and (2) a modification of the two processing principles I had suggested earlier:

1. *The prominence of the prosodic high attachment cue (pause):*

   Assuming acoustic features of a pause (such as pitch drop at and lengthening of the pre-pause word, duration of silence) were being perceived as more or less prominent, this perceived difference in prominence might impact processing of a prosodically supported syntactic structure. Perceived prominence is dependent on the prosodic cue itself and on the perceivers' acoustic sensitivity.

2. *The surprisal associated with the high attachment cues (pause v. high attachment):*

   Assuming there is a default structural interpretation for a given structure, this principle marks a cue as surprising and thus more salient when it is not in line with the preferred default. Assuming surprisal exists, a cue that is more surprising and salient carries greater weight and will exert more influence during processing.

Following principle number 1, if the pause is prominent enough it is noted as surprising and therefore carries weight in terms of subsequent processing. If it is not prominent enough its surprisal value is lessened and the cue would carry less weight for subsequent processing.

3. *The type of structural revision necessary:*

   Assuming structural revision is caused by a mismatch between different cues, the difference in weight will determine the type of structural revision. That is, an early surprising and therefore weightier cue can overrule a later default-supporting cue; and an early default-supporting cue can be suppressed by a later surprising cue.

   However, an early surprising cue such as a pause could overrule a later default-supporting cue such as semantic LA **only if** the pause was perceived as sufficiently prominent (see principle 1). An early default-supporting cue such as a no-pause could be overruled by a later surprising cue such as semantic HA **only if** the semantic HA cue adds enough weight in the time available for revision.

The first principle takes into account the differential priming results. I found priming effects for a prominent prosodic manipulation across all participants. With a subtler prosodic manipulation two clusters were found that showed opposing priming results. That is, the notion of surprisal at least in terms of prosody and the severity of the structural revision seem to depend on either the subjects' acoustic discrimination ability or the acoustic prominence of the pause cue itself. The strongest evidence in support of this is firstly the observation that no detectable individual differences were found in the pilot study that used a more prominent acoustic pause cue.

Secondly, the repeated discovery of the two participant clusters with different acoustic discrimination abilities argues for the prominence principle to be taken into account.

In general, the *surprisal* principle assumes the existence of a default and preferred interpretation for any given syntactic structure. In English, low attachment of the RC seems preferred and, consequently, the default. A cue that is not in line with the default such as prosodic support for HA (pause) or semantic support for HA might be perceived as surprising (Jaeger & Snider, 2008; Scheepers, 2003) and therefore more salient.

It is further assumed that surprisal leads to an increase in influence that the surprising information exerts during processing. Consequently, a more surprising, salient and important cue applies its weight during online processing and might still influence consecutive processing as evidenced in the target completions after priming.

Thus, in the case of cue conflict (*pause + semantics LA* and *no-pause + semantics HA*) the more surprising cue (*pause* v. *semantic HA*) exerts influence during online processing and primes subsequent HA target completions.

To be more specific, when processing the main clause before the relative clause (*Someone shot the servant of the actress …*) the syntactic structure that is most compatible with the input is built up.

Whenever a pause is encountered before the RC, and is perceived as prominent (*prominence principle*), it is a salient and surprising (*surprisal*) cue to high attachment of the RC. During subsequent processing, this prominent and surprising information leads to revision of the preferred low attachment structure and a high attachment structure is build up. High attachment is generally dispreferred and might be more difficult to process as the intended attachment host (*the servant*) has been encountered earlier than the preferred attachment host (*the actress*). Thus when encountering the last word, processing difficulty might be expected by the parser. The processing difficulty could be due to the generally dispreferred nature of the high attachment structure that was cued by the surprising pause. This is independent of disambiguating semantic information, and the interaction of prosody* region in the ERP experiment is an indication of this. This early N400 effect for pause compared to no-pause conditions lends support to the assumption that a prosodically cued HA structure is dispreferred and induces processing difficulty.

On the other hand, when no-pause was present in the speech signal a low attachment structure is followed because of its preferred status. At the last word, surprising semantic information in support of high attachment might be encountered. High attachment induces online processing difficulty because of its dispreferred nature. The early negativity observed at the last word for the high compared to low attachment plausibility cue accentuates this.

To conclude, the surprisal principle that was suggested by the priming data finds support in the ERP data. Note, however, that the surprisal principle should only exert influence if it has been perceived as prominent and has therefore been recognised as surprising. This conclusion is based on the discovery of the two clusters of participants that differ in acoustic ability.

I have pointed out before that surprise as a sole processing principle cannot account for the observed results. The strongest evidence is derived from the lack of a priming effect in the pause and semantic HA condition where both primes are not in line with the general LA default attachment.

Reliable priming effects were found for conditions where the cues supported different attachment decisions. This is further supported by the ERP results that revealed processing difficulty for mismatching cues as compared to matching cues.

Two different conflict situations were present in the materials i.e., *pause + semantic LA* and *no-pause + semantic HA*.

In the first case (*pause + semantic LA*), a surprising pause is presented *early* just before the relative clause, which should cue a high attachment structure to be built. A default, low attachment plausibility cue is encountered *later* and it should falsify the current HA structure. However, to do that the positive and weighty evidence of the pause would need to be ignored. As Pauker et al. (2011) argued, it might be more costly and possibly take longer to undo the positive and highly salient evidence as encountered in a pause. Thus, the HA structure initiated by the pause cue is very strongly activated during processing of the prime sentence and it therefore encourages high attachment in the target completions. Thus, an early surprising and therefore weightier cue can overrule a later default-supporting cue. This is certainly true for priming and might be understood in terms of processing difficulty during online processing.

A late positivity is observed for this type of conflict in the second recording. This suggests that participants have learned to recognise and to try and solve the cue conflict. The resolution of the cue conflict might be understood as an attempt to undo the very prominent and weighty pause cue. The processing difficulties seemed to start early on (between 300-500ms) and it was found to last for a substantial amount of time (500-900ms).

Conversely, in the second case of cue conflict (*no-pause + semantic HA*) *early* prosodic evidence is in line with the low attachment default. *Later* plausibility information disambiguates the current LA structure towards high attachment.

Since no strong prosodic cue was presented early on, the later surprising cue can overrule the default-supporting one and lead to high attachment target completions. That is, revision ensues and is solved towards high attachment because the semantic information was surprising. A late positivity effect was also observed for this type of conflict. It was found in the same time windows as the other type of conflict, which indicates that either type of structural revision necessary induces processing difficulty.

No difference in terms of ERPs were found between the types of structural revision. This is not in accordance with the results reported by Pauker et al. (2011) who found different effects for missing as opposed to misplaced boundaries. Thus, it might be premature to conclude that one is more difficult than the other. Further research is warranted.

Again it is important to emphasise that the type of revision only comes into play when the dispreferred high attachment cue has actually been perceived as prominent and surprising.

To conclude, I have discussed the processing principles that I have suggested in a way that highlights their internal consistency. However, embedding the principles into the wider theoretical debate is possible.

It is entirely imaginable that processing heuristics such as Frazier's *Late closure* and *Minimal attachment* guide processing of the RC attachment in two stages. The suggested principles are compatible with such an assumption.

However, to align the here observed ERP effects with a syntax-first, two-stage model such as Friederici's Neurocognitive model of auditory sentence processing (Friederici, 2002, 2011) might turn out to be very difficult.

An early negativity was found for high compared to low attachment conditions in the present ERP experiment. If this negativity reflected difficulty due to initial structure building I would expect a word category violation. This is not the case here. All items were syntactically well-formed, i.e. no unexpected word category occurred at this position. Thus, the negativity should not be interpreted as an ELAN. Equally, the anterior negativity should not be interpreted as a LAN effect since no morphosyntactic violation was present.

The early negativity was found for pause compared to no-pause conditions. It is difficult to explain what might cause context integration problems when assuming Friederici's model as a framework.

The only effect that might be compatible with this model is the observed anterior positivity. Firstly, a mismatch between a syntactic structure and incoming disambiguating information should result in a posterior P600. However, the disambiguation is not syntactic in nature, thus the difference in topography could be incorporated into the model, especially since previous evidence supports this finding (e.g. Kaan & Swaab, 2003). Secondly, Friederici assumes that ambiguous structures are processed slightly differently. Therefore, a different P600 as a response to the resolution of an ambiguity might be conceivable. The anterior positivity could reflect this but it would need to be further verified should it be included into this model.

To relate the presently observed results to parallel, constraint-satisfaction models could be challenging as well. It is tempting to translate the idea of the proposed processing principles into constraints that apply weight in a competition situation. That is, surprisal associated with unexpected and possibly dispreferred attachment cues might turn the competition between different attachment options in favour of high attachment.

However, parallel, constraint-satisfaction models usually do not include the notion of revision and would therefore struggle to incorporate the last principle.

Thus, more theoretical work needs to be done to incorporate the results reported here into existing theoretic frameworks such as Hagoort's MUC model (Hagoort, 2003; 2005).

Future research and concluding remarks

The body of work presented here is the first to investigate the interaction of overt prosodic and plausibility cues to relative clause attachment. The results are complex but very exciting at the same time.

In terms of research methodology, I could show that priming can be used to investigate subtle prosodic manipulations. This is very encouraging since research into prosody has somewhat suffered from not being able to show strong effects.

Additionally, my work underlines the fact that direct replications should be an essential part of research work being done. How else could we show that an effect is true? This also emphasises the importance of carefully interpreting initial results such as the observed ERP effects. Further research should investigate whether the effects reported here are replicable and how much validity the interpretation that I have provided actually holds.

However, at the same time it has become apparent that much rests on the materials being used. Here materials with prosodic manipulations of different strength were used. Variability of material is something that future research might look at. That is for example, what are the essential acoustic features to elicit a closure positive shift (CPS) as an ERP response? What acoustic features determine when a prosodic cue is “prominent enough” to influence processing across all participants and when is it “subtle enough” to tap into individual differences of acoustic discrimination ability?

One of the major findings of this thesis is the fact that individual differences such as participants’ acoustic abilities could influence the use of prosodic cues during language processing. It might be beneficial for future research into language processing to take individual differences into account. Opposing results reported in the literature might be related to unrecognised differences between participants used in the studies.

Also, models of language processing could profit by allowing differences to manifest. After all, theories of language processing try to account for language processing in general by pointing out what is common in language users. But to look at commonalities might be just one way of studying language processing. Much can be gained from considering in what way people vary. As observed here, commonality can emerge when variability is kept in mind.

I will leave the last remarks to Ben Goldacre whose words brilliantly summarise the work presented here:

Appendix 1 Items and target fragments used for the pilot study

Item	Prime sentence	Target fragment
1	The criminal shot the servant of the actress who was serving tea.	The tourist guide mentioned the bells of the church that ____ .
	The criminal shot the servant of the actress who was very famous.	
2	The student thought about the content of the book that was very abstract.	The manager waited for the musicians of the pop star who ____ .
	The student thought about the content of the book that was rather heavy.	
3	Today Jane spoke with the father of the pupil who owned a shop.	The commission referred to the source of the donations that ____ .
	Today Jane spoke with the father of the pupil who was doing well in class.	
4	The mechanic repaired the engine of the race car that had titanium pistons.	Nora visited the students of the piano teacher who ____ .
	The mechanic repaired the engine of the race car that had a new kind of spoiler.	
5	The politician referred to the source of the information that had contacted him.	The chauffeur met the representative of the state guests who ____ .
	The politician referred to the source of the information that was not newsworthy.	
6	Mary babysits the child of the musician that was in the cot next-door.	The tutor advised the students of the lecturer who ____ .
	Mary babysits the child of the musician that had a beard.	
7	I know the father of the secretary who is a retired policeman.	The superintendent checked the earnings of the company that ____ .
	I know the father of the secretary who married a doctor.	
8	John detests the wife of the artist who is pregnant	The bus driver talked to the leader of the boy scouts who ____ .
	John detests the wife of the artist who wore a mustache.	
9	Someone smashed the window of the car that made of tinted glass.	The farmhand fed the calves of the cow that ____ .
	Someone smashed the window of the car that had a big exhaust.	
10	Paddy showed the costumer the mother of the puppy that was old.	The pensioner complained about the content of the fliers that ____ .
	Paddy showed the costumer the mother of the puppy that was newborn.	
11	Peter approached the manager of the pop star who formulated the contract.	The frost ruined the harvest of the fruit farms that ____ .
	Peter approached the manager of the pop star who released a new album.	
12	The analyst commented on the development of the market that was surprising.	John met the supervisor of the employees who ____ .
	The analyst commented on the development of the market that was growing.	
13	Eileen liked the colour of the dress that was bright and fresh.	The social worker greeted the nurse of the senior-citizens who ____ .
	Eileen liked the colour of the dress that was made of silk.	
14	Daniela was very happy about the funding of the project that will be sufficient to pay the subjects.	We were amused at the articles of the newspaper that ____ .
	Daniela was very happy about the funding of the project that will be conducted within the department.	
15	The fans admired the coach of the wrestler who trained him for years.	The insurance company covered the furniture of the apartments

	The fans admired the coach of the wrestler who injured his knee.	that ____ .
16	The board discussed the summary of the survey that highlighted the most important points.	The hacker attacked the web sites of the service provider that ____ .
	The board discussed the summary of the survey that was undertaken in the previous year.	
17	The scientist is happy about the result of the experiment that clearly confirmed her prediction.	A stranger blackmailed the butler of the royals who ____
	The scientist is happy about the result of the experiment that uses a new kind of method.	
18	John argued with the brother of the girl who was a sales manager.	The scientist criticised the method of the studies that ____ .
	John argued with the brother of the girl who was a bully at school.	
19	James knocked on the door of the house that was left ajar.	The secret service confiscated all files of the organisation that ____ .
	James knocked on the door of the house that had a new roof.	
20	The salesperson ignored the mother of the baby who was searching her handbag.	The assassin saw the bodyguard of the diplomats who ____ .
	The salesperson ignored the mother of the baby who was making tantrums.	
21	The advisor commented on the progress of the work that appeared slower then expected.	The astronomer observed the stars of the spiral galaxy that____ .
	The advisor commented on the progress of the work that had been carried out recently.	
22	The PR manager looked at the advert of the company that appeared on the screen in front.	The homeowner kept the letters of the estate agency that ____ .
	The PR manager looked at the advert of the company that owned a big production studio.	
23	The vet examined the leg of the horse that appeared to be broken.	The porter smiled at the children of the hotel resident who ____ .
	The veterinarian examined the leg of the horse that was supposed to win the race.	
24	The journalist interviewed the agent of the movie star who had made the contract.	The scholar studied the language of the tribes that ____ .
	LA: The journalist interviewed the agent of the movie star who had won an Oscar.	

Appendix 2 Items and target fragments used for main priming experiment 1,2

Items used for the ERP experiment

Item number	Sentence	Target
1	Mandy painted the front door of the flat[] that had a new cat flap.	On YouTube, we saw the teacher of the teenagers who ...
	Many painted the front door of the flat that had a new tenant.	
2	The protester shouted at the manager of the bank [] that had received a bonus.	Everyone laughed at the pictures from the Christmas party that...
	The protester shouted at the manager of the bank that had received a subsidy.	
3	The community listened to the rep of the company [] that was handsome.	Harry sold the tomatoes of the farmer that
	The community listened to the rep of the company that was bankrupt.	
4	Mia interrupted the secretary of the mayor [] who had been promoted.	A poster announced the exhibition of the paintings that ...
	Mia interrupted the secretary of the mayor who had been elected.	
5	Andrew talked to the tutor of the student [ ] who had graded the exam.	The keepers talked about the enclosure of the sea lions that...
	Andrew talked to the tutor of the student who had taken the exam.	
6	Peter descaled the fence of the allotment [] that was very rusty.	The prosecutor talked to the lawyer of the defendants who...
	Peter descaled the fence of the allotment that was very tidy.	
7	The children laughed about the clown in the arena [] that had been singing.	The reporter interviewed brothers of the girl who...
	The children laughed about the clown in the arena that had been empty.	
8	Today we met the professor of the PhD student [] who had recently retired.	The anthropologist listened to the screams of the monkey that ...
	Today we met the professor of the PhD student who had recently enrolled.	
9	The teacher welcomed the father of the youngster [ ] who was a sales rep.	The scientist checked the equipment of the labs that...
	The teacher welcomed the father of the youngster who was a boy scout.	
10	Erin analysed the data of the experiment [ ] that she had collected.	The drought ruined the crop of the plantations that ...
	Erin analysed the data of the experiment that she had conducted.	
11	The documentary portrayed the relatives of the flood survivor [ ] who had been searching.	The board discussed the outcome of the exams which ...
	The documentary portrayed the relatives of the flood survivor who had been rescued.	
12	The gardener sold the bulbs of the flowers [ ] that had been firm.	The bus driver talked to the leader of the boy scouts who ...
	The gardener sold the bulbs of the flowers that had been cut.	

13	The police took fingerprints from the window of the flat [ ] that had been repaired.	The IT administrator updated the software of the computers that ...
	The police took fingerprints from the window of the flat that had been rented.	
14	Karen admired the author of the book [ ] that was laughing.	The head teacher recognized the mother of the twins who ...
	Karen admired the author of the book that was launched.	
15	The smith talked to the squire of the night who was very servile.	The delegation prepared for the last day of the meetings that...
	The smith talked to the squire of the night who was very noble.	
16	The nanny complained about the teenage daughter of the professor [ ] who was very spoiled.	The porter apologised for the cancellation of the trains which ...
	The nanny complained about the teenage daughter of the professor who was very aged.	
17	Ben double checked the opening hours of the agency [ ] that had changed.	The magazine portrayed the sons of the fashion designer who ...
	Ben double checked the opening hours of the agency that had moved.	
18	The journalist interview the husband of the actress [ ] who was balding.	The commission referred to the source of the donations that ...
	The journalist interviewed the husband of the actress who was pregnant.	
19	The editor laughed about the spelling of the article [ ] that was wrong.	The presenter announced the conductor of the musicians who ...
	The editor laughed about the spelling of the article that was long.	
20	The gladiator looked at the slave of the consul [ ] who was to be freed.	The nurse saw the friends of the patient who...
	The gladiator looked at the slave of the consul who was to be stabbed.	
21	The company paid for the gold from the mine [ ] that was stolen.	The limousine stopped by the groupies of the pop star who ...
	The company paid for the gold from the mine that was opened.	
22	Maria wrote the invitations for the party [ ] that she wanted to give.	The documentary was about the bridges of the river that...
	Maria wrote the invitations for the party that she wanted to post.	
23	At the party he met the sister of the inventor [ ] who was wearing a cocktail dress.	The chef read the menu of the banquet that ...
	At the party he met the sister of the inventor who was wearing a bowler hat.	
24	The singer practised the lyrics of the song [ ] that had many difficult words.	The talent scout talked to the manager of the football players who ...
	The singer practised the lyrics of the song that had many difficult notes.	
25	The government condemned the violation of the ceasefire [ ] that had been provoked.	The security guards searched the advisors of the president who...
	The government condemned the violation of the ceasefire that had been declared.	
26	Matthew interviewed the attorney of the victim [ ] who had given legal advise.	A mechanic repaired the tyres of the sports car that ...
	Matthew interviewed the attorney of the victim that had given legal evidence.	
27	The old man wrote down the legend of the ghost ship [ ] that he had told many times	The trainee nurse met the patients of the doctor who...

	The old man wrote down the legend of the ghost ship that he had seen many times.	
28	The pope visited the monks of the monastery [ ] that had been ordained.	The pensioner complained about the font size of the brochures that...
	The pope visited the monks of the monastery that had been bombed.	
29	The surveyor commented on the garden of the house [ ] that needed a new fence.	The solicitor summoned the heirs of the baron who ...
	The surveyor commented on the garden of the house that needed a new roof.	
30	The vet looked at the teeth of the dog [ ] that had been extracted	The terrorists broadcasted a video of their hostages that ...
	The vet looked at the teeth of the dog that had been sedated.	
31	The article mentioned the management of the Aid programme [ ] that was about to resign.	The House of Commons started the debate on the expenses which ...
	The article mentioned the management of the Aid programme that was about to commence.	
32	The vicar welcomed the aunt of the groom [ ] who is a midwife.	Peter admired the trucks of the fire brigade that...
	The vicar welcomed the aunt of the groom who is a boxer.	
33	Sergeant Wood commanded the guards of the prisoners [ ] who were on a long term contract.	On his door step, Johnny found the kittens of the cat that ...
	Sergeant Wood commanded the guards of the prisoners who were on a long term sentence.	
34	Harry was thrilled about the character of the story [ ] that had been killed.	Britney collected the eggs of the chicken that...
	Harry was thrilled about the characters of the story that had been true.	
35	The chauffeur picked up the groupie of the band [ ] that was blond.	The chairman explained the recommendations of the commission that...
	The chauffeur picked up the groupie of the band that was loud.	
36	Monika wrote a book about the memories of the soldiers [ ] that were vivid.	The union targeted the workers of the planter who ...
	Monika wrote a book about the memories of the soldiers that were wounded.	
37	The punter talked to the pimp of the prostitute [ ] who was brutal	The internal report named the costumers of the sales assistant who ...
	The punter talked to the pimp of the prostitute who was pretty.	
38	He had witnessed the flood on the island [ ] that had been a disaster.	The psychologist contacted the social worker of the junkies who ...
	He witnessed the flood of the island that had been a paradise.	
39	The parties did not comment on the breakdown of the negotiations [ ] that had been sudden.	The historian mentioned the knights of the king who...
	The parties did not comment on the breakdown of the negotiations that had been hostile.	
40	The reporter tracked down the maid of the millionaire [ ] who had been dismissed.	The custodian locked the doors of the mansion that ...
	The reporter tracked down the maid of the millionaire who had been kidnapped.	
41	The Sun featured the bride of the prince [ ] who had been a florist.	The costumer complained about the prices in the catalogue that...
	The Sun featured the bride of the prince who had been a macho.	

42	The gardener examined the fruit of the tree [ ] that was ripe.	Later that day, Betty saw the grandchildren of her neighbour who…
	The gardener examined the fruit of the tree that was tall.	
43	The court heard the social worker of the addict [ ] who was a professional.	The anthropologist listened to the screams of the monkey that …
	The court heard the social worker of the addict who was a criminal.	
44	The food critic complained about the quality of the steak [ ] that was low.	The bishop talked to the prior of the monks who…
	The food critic complained about the quality of the steak that was dry.	
45	John read about the painting of the village [ ] that had been stolen.	The document referred to the ancestors of the king who…
	John read about the painting of the village that had been flooded.	
46	The tourists were concerned about the guard of the gate [ ] that had been puking.	Simon visited the students of the voice coach who …
	The tourists were concerned about the guard of the gate that had been unlocked.	
47	The artist finished the graffiti on the wall [ ] that was colourful.	The cameraman filmed the lambs of the ewe that…
	The artist finished the graffiti on the wall that was instable.	
48	The editor commented on the progress of the book [ ] that was very slow.	Ben hated the overseer of the slaves who …
	The editor commented on the progress of the book that was very short.	
49	The mechanic looked after the tires of the race car [ ] that needed more air.	The couple thought about the offers of the letting agency that
	The mechanic looked after the tires of the race care that needed more fuel.	
50	The tourist guide mentioned the cemetery of the village [ ] that was overgrown.	The workers cut down the trees of the avenue that …
	The tourist guide mentioned cemetery of the village that was deserted.	
51	The commission presented the facts about the disease [ ] that had been gathered.	The police arrested the accomplice of the gangsters who…
	The commission presented the facts about the disease that had been deadly.	
52	The vet weighed the calf of the cow [] that had been fed.	The paparazzi spotted the hairdresser of the celebrities who…
	The vet weighed the calf of the cow that had been milked.	
53	Paddy inspected the cage of the tiger [ ] that had been cleaned.	The caretaker cleaned the tank of the dolphins that …
	Paddy inspected the cage of the tiger that had been caught.	
54	The entire household looked for the key of the gate [ ] that had been lost.	The couple thought about the offers of the letting agency that
	The entire household looked for the key of the gate that had been locked.	
55	The public was upset about the comment on the programme [ ] that had been outrages.	The doctor visited the parents of the girl who …
	The public was upset about the comment on the programme that had been broadcasted.	
56	The brochure showed the prize of the mansion [ ] that was astronomical.	On X-Factor, Dermot teased the mentor of the contestants who …
	The brochure showed the prize of the mansion that was neoclassical.	

57	The paramedics tended to the victim of the car crash [ ] that had been injured.	The villagers were concerned about the approach of the troops that…
	The paramedics tended to the victim of the car crash that had been severe.	
58	The witness described the scene of the crime [ ] that had been sealed off.	The security man double-screened the luggage of the passengers that …
	The witness described the scene of the crime that had been vicious.	
59	The midwife congratulated the father of the new born [ ] who was very proud.	Sally looked at the windows of the house that …
	The midwife congratulated the father of the new-born who was very weak.	
60	The scientist praised the publication of the study [ ] that had a lot of figures.	The nurse greeted the head of the paramedics who …
	The scientist praised the publication of the study that had a lot of subjects.	
61	The frost ruined the harvest of the farm [ ] that was due to commence.	On YouTube, we saw the teacher of the teenagers who …
	The frost ruined the harvest of the farm that was due to expand.	
62	Kurt bought the postcard of the town [ ] that he wanted to send.	The prosecutor talked to the lawyer of the defendants who…
	Kurt bought the postcard of the town that he wanted to leave.	
63	The undercover agent tailed the dealer of the celebrity [ ] who had been on probation.	The keepers talked about the enclosure of the sea lions that…
	The undercover agent tailed the dealer of the celebrity who had been on Big Brother.	
64	James knocked on the door of the house [ ] that was left ajar.	The nurse saw the friends of the patient who…
	James knocked on the door of the house that was left vacant.	
65	The salesperson served the mother of the toddler [ ] who was making a purchase.	Harry sold the tomatoes of the farmer…
	The salesperson served the mother of the toddler who was making a tantrum.	
66	The assassin saw the bodyguard of the diplomat [ ] who was carrying a hand gun.	Everyone laughed at the pictures from the Christmas party that…
	The assassin saw the bodyguard of the diplomat who was carrying a brief case.	
67	The advisor commented on the progress of the work [ ] that had been steady.	The nurse greeted the head of the paramedics who …
	The advisor commented on the progress of the work that had been tricky.	
68	Jason kept the letters of the estate agency [ ] that he had recently received.	The security guards searched the advisors of the president who…
	Jason kept the letters of the estate agency that he had recently hired.	
69	We witnessed the fire of the cottage [ ] that was spreading.	John distrusted the apprentices of the plumber who …
	We witnessed the fire of the cottage that was shabby.	
70	The biologist looked at the membranes of the cells [ ] that were unusually thin.	The union targeted the workers of the planter who …
	The biologist looked at the membranes of the cells that were unusually large.	
71	Nelly slapped the mistress of the pirate [ ] who wore a black dress.	The scientist checked the equipment of the labs that…

	Nelly slapped the mistress of the pirate who wore a black beard.	
72	Darleen looked at the man in the business suit [ ] that was very moody.	The board discussed the outcome of the exams which ...
	Darleen looked at the man in the business suit that was very pricey.	
73	Chris knew the interest rate of the bank account [ ] that he wanted to earn.	The bus driver talked to the leader of the boy scouts who ...
	Chris knew the interest rate of the bank account that he wanted to close.	
74	The tale was about the tailor of the king [ ] who was very skilled.	The IT administrator updated the software of the computers that ...
	The tale was about the tailor of the king who was very rich.	
75	The foster mother looked at the child in the room [ ] that was orphaned.	The trainee nurse met the patients of the doctor who...
	The foster mother looked at the child in the room [ ] that was soundproof.	
76	The journal mentioned the discovery of the new species [ ] that had been published.	On X-Factor, Dermot teased the mentor of the contestants who ...
	The journal mentioned the discovery of the new species that had been captured.	
77	The conductor handed out the new repertoire of the choir [ ] that he had compiled.	The historian mentioned the knights of the king who...
	The conductor handed out the new repertoire of the choir that he had founded.	
78	The documentary featured the victim of the murderer [ ] who had been cremated.	The pensioner complained about the font size of the brochures that...
	The documentary featured the victim of the murderer who had been arrested.	
79	The guild was concerned about the apprentice of the baker [ ] who didn't want to learn.	The documentary was about the bridges of the river that...
	The guild was concerned about the apprentice of the baker who didn't want to teach.	
80	The bible never mentioned the daughter of the prophet [ ] who had been exceptionally gorgeous.	The drought ruined the crop of the plantations that ...
	The bible never mentioned the daughter of the prophet who had been exceptionally righteous.	
81	Everybody knew about the wife of the warrior [ ] who was said to be very beautiful.	A poster announced the exhibition of the paintings that ...
	Everybody knew about the wife of the warrior who was said to be very battle scarred.	
82	The agency contacted the referee of the applicant [ ] who was a University professor.	The chef read the menu of the banquet that ...
	The agency contacted the referee of the applicant who was a university graduate.	
83	"Karate Kid" is about the student of a Kung Fu Master who was a teenager.	The caretaker cleaned the tank of the dolphins that ...
	"Karate Kid" is about the student of a Kung Fu Master who was a veteran.	
84	The comic described the attack of the aliens [ ] that had been fatal.	The talent scout talked to the manager of the football players who ...
	The comic described the attack of the aliens that had been breeding.	
85	The organiser thought about the date of the auction [ ] that he wanted to set.	The head teacher recognised the mother of the twins who...
	The organiser thought about the date of the auction that he wanted to run.	

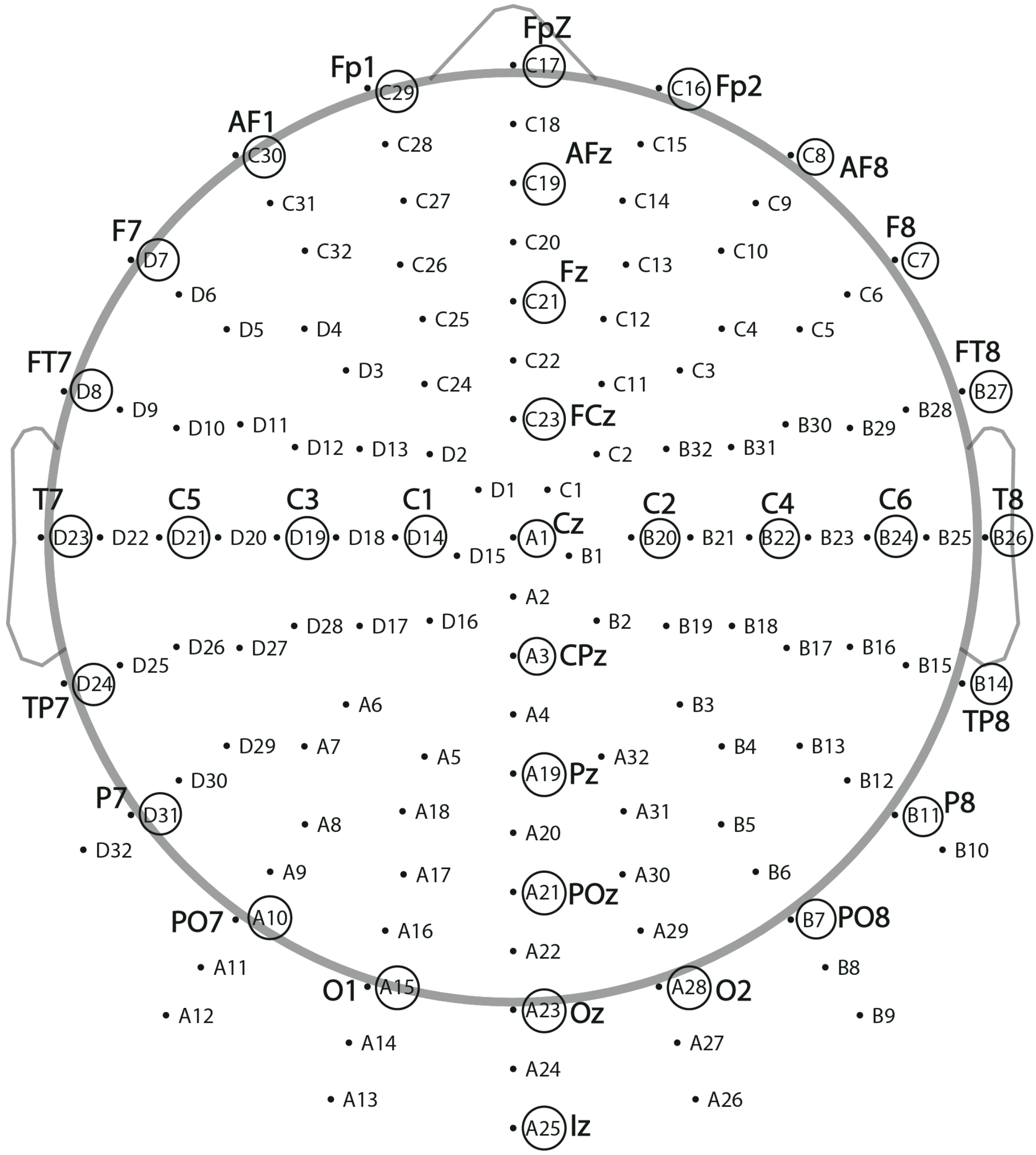
86	The thesis acknowledged the supervisor of the author [ ] who had been very helpful. The thesis acknowledged the supervisor of the author who had been very grateful.	A mechanic repaired the tyres of the sports car that …
87	Silvia comforted the widow of the Mafioso [ ] who was upset. Silvia comforted the widow of the Mafioso who was killed.	The delegation prepared for the last day of the meetings that…
88	Benedict describes the foes of the early Christians [ ] who were mostly heathens. Benedict describes the foes of the early Christians who were mostly martyrs.	The commission referred to the source of the donations that …
89	The clinic contacted the counsellor of the technician [ ] who was an expert in psychology. The clinic contacted the counselor of the technician who was an expert in engineering.	Sally looked at the windows of the house that …
90	Inspector Hardy questioned the kidnapper of the millionaire [ ] who had been caught. Inspector Hardy questioned the kidnapper of the millionaire who had been freed.	The minister ignored the suggestions of the committee that…
91	The journalist interrupted the spokeswoman of the minister [ ] who was faltering. The journalist interrupted the spokeswoman of the minister who was detained.	The terrorists broadcasted a video of their hostages that …
92	The researcher was dreading the deadline of the application [ ] which could not be extended.. The researcher was dreading the deadline of the application which could not be submitted.	Later that day, Betty saw the grandchildren of her neighbour who…
93	The audience awaited the launch of the ship [ ] that was about to take place. The audience awaited the launch of the ship that was about to sail off.	On his door step, Johnny found the kittens of the cat that …
94	The tourist guide warned against the path to the hotel [ ] that was muddy. The tourist guide warned against the path to the hotel that was pricey.	Simon visited the students of the voice coach who …
95	The hospice contacted the doctor of the patient [ ] who had been practising for years. The hospice contacted the doctor of the patient who had been suffering for years.	The House of Commons started the debate on the expenses which …
96	Eric first bought the food for the party [ ] that he wanted to cook. Eric first bought the food for the party that he wanted to throw.	The presenter announced the conductor of the musicians who …
97	The head teacher awaited the money for the classroom [ ] that was to be transferred. The head teacher awaited the money for the classroom that was to be painted.	The psychologist contacted the social worker of the junkies who …
98	Ray packed the nuts for the squirrels [ ] that he had picked in the park the other day. Ray packed the nuts for the squirrels that he had fed in the park the other day.	The limousine stopped by the groupies of the pop star who …
99	The reporter bribed the dentist of the pop star [ ] who had his own dental practice. The reporter bribed the dentist of the pop star who had his own record label.	Britney collected the eggs of the chicken that…
100	The crowed ignored the opponent of the winner who was really frustrated.	The cameraman filmed the lambs of the ewe that…

	The crowed ignored the opponent of the winner who was really ecstatic.	
101	The advert featured the personal coach of the runner [ ] who was very strict.	The costumer complained about the prizes in the catalogue that…
	The advert featured the personal coach of the runner who was very fast.	
102	The painting showed the chaplain of the prince [ ] who later became pope.	Peter admired the trucks of the fire brigade that…
	The painting showed the chaplain of the prince who later became king.	
103	The community mourned the child of the miner who had been very small.	The anthropologist listened to the screams of the monkey that …
	The community mourned the child of the miner who had been very sad.	
104	The story mentioned the girl friend of the sailor [ ] who looked like Lady Di.	The couple thought about the offers of the letting agency that
	The story mentioned the girl friend of the sailor who looked like Gordon Brown.	
105	The KGB kidnapped the granny of the spy [ ] who played a lot of bingo.	The custodian locked the doors of the mansion that …
	The KGB kidnapped the granny of the spy who played a lot of poker.	
106	Polly read the description of the steam boat [ ] that was really technical.	The paparazzi spotted the hairdresser of the celebrities who…
	Polly read the description of the steam boat that was really comfortable.	
107	The children watched the birth of the foal [ ] that was difficult.	The document referred to the ancestors of the king who…
	The children watched the birth of the foal that was dappled.	
108	The board published the losses of the company [ ] that had been substantial.	Ben hated the overseer of the slaves who …
	The board published the losses of the company that had been registered.	
109	Stacy visited the homepage of the MSP [ ] that had been redesigned.	The chairman explained the recommendations of the commission that…
	Stacy visited the homepage of the MSP that had been elected.	
110	Herbert bought the book about the war [ ] that had a lot of pictures.	The police arrested the accomplice of the gangsters who…
	Herbert bought the book about the war that had a lot of victims.	
111	The activist described the clearing of the rain forest [ ] that has been on-going.	The magazine portrayed the sons of the fashion designer who …
	The activist described the clearing of the rain forest that has been recovering.	
112	The economist stressed the analysis of the market [ ] that was accurate.	The villagers were concerned about the approach of the troops that…
	The economist stressed the analysis of the market that was expanding.	
113	The CCTV captured the violence of the group [ ] that was taking place.	The internal report named the costumers of the sales assistant who …
	The CCTV captured the violence of the group that was taking drugs.	
114	The staff celebrated the review of the restaurant [ ] that was well written.	The bishop talked to the prior of the monks who…
	The staff celebrated the review of the restaurant that was well managed.	

115	Dan finished the download of the file [ ] that had been fast. Dan finished the download of the file that had been small.	The doctor visited the parents of the girl who ...
116	Stan dated the niece of the governor [ ] who was a prostitute. Stan dated the niece of the governor who was an paedophile.	The workers cut down the trees of the avenue that ...
117	The janitor pointed at the organ of the concert hall [ ] that had just been installed. The janitor pointed at the organ of the concert hall that had just been restored.	The solicitor summoned the heirs of the baron who ...
118	Last night we went to the theatre by the lake [ ] that had been burnt. Last night we went to the theatre by the lake that had been drained.	The security man double-screened the luggage of the passengers that ...
119	We were interested in the dishes of the Greek islands [ ] that were really delicious. We were interested in the dishes of the Greek islands that were really romantic.	The porter apologised for the cancellation of the trains which ...
120	Narnia is about the secret of the wardrobe [ ] that some children had revealed. Narnia is about the secret of the wardrobe that some children had opened.	The reporter interviewed brothers of the girl who...

Appendix 3 BIOSEMI electrodes layout

BIOSEMI 128 electrodes locations



Appendix 4 Summary of Augurzky (2005)

Augurzky (2005) looked at RC attachment in German, a high attachment preference language. She investigated different aspects of RC attachment. I will concentrate on two of her experiments.

In a reading experiment, she used complex noun phrases that involved a preposition (*von*, 2 a – d) or Genitive case marking (1 a – d). The RC attachment was disambiguated via gender agreement between relative pronoun and attachment host. Thus disambiguation took place early at the relative pronoun.

At this early point of disambiguation, a broadly distributed negativity which resembled a N400 was found when attachment was forced high compared to low. This negativity was independent of the preposition (*von*) and Genitive case marking in the complex noun phrase. Thus, processing difficulty was found for the preferred high attachment compared to the dispreferred low attachment. However, processing difficulty might not necessarily be due to a preference. It could also be due to the increased structural complexity of HA and it might pose higher demands on general cognitive resources.

In another EEG experiment the same sentences were used but adopted for the auditory domain. Low attachment was assumed to be cued by a pause after NP1 which groups NP2 and the RC in one phrase. High attachment was cued by a pause after NP2 which prevents local low attachment and forces the parser to look for another attachment host. Prosody was crossed with syntactically determined attachment via case and gender marking at the relative pronoun.

That is, the design comprised three factors: *prosody* (pause after NP1 v. pause after NP2), *preposition* within the complex noun phrase (*von* v. Genitive case marking) and *attachment* at the relative pronoun (high v. low).

High attachment prosody (pause after NP2 before the RC) elicited one CPS but, surprisingly, low attachment prosody (pause after NP 1) elicited two CPSs. The first appeared at the NP1 boundary position and the second one at the NP2 position.[17]

[17] This was perhaps only a little surprising. The acoustic measurements at NP2 showed a difference in pre-final lengthening between high and low attachment prosody at this point. But pitch movement across the NP2 for both prosodies was highly comparable. Thus, listeners might have perceived two acoustic

When Genitive case marking was used within the complex noun phrase, a pause after NP2 (high attachment prosody) did not seem to influence processing, i.e. no differences between conditions were observed. Low attachment prosody (pause after NP1) on the other hand, seemed to have an immediate impact on processing. A N400-like negativity was observed for RCs that were forced to attach high.[18]

When "von" prepositions linked the two nouns in the complex noun phrase, prosody had a different impact. High attachment prosody (pause after NP2) triggered an N400 response after attachment was syntactically disambiguated high compared to low. Low attachment prosody (pause after NP1) did not show any ERP responses when comparing syntactic high to low attachment.

To summarise, Augurzky's (2005) results sketch a somewhat confusing picture of RC attachment in German. Early disambiguation at the relative pronoun, resulted in a broadly distributed negativity for syntactic high attachment regardless of the type of complex noun phrase.

During listening, the prosodic manipulation and syntactic disambiguation interacted differently for complex noun phrases that involved the preposition *von* and the Genitive case marking. However, the differential results were to some extent difficult to interpret since the auditory materials might have included an acoustic confound in the low attachment prosody conditions.

---

boundaries for low attachment prosody, which could have implications for processing further down the line.

[18] If the 2 CPS' observed in this condition (see previous footnote) are a reflection of structuring of the input then an N400 effect might be linked to the presence of these two acoustic cues. It could be interpreted along the lines of Steinhauer et al. (1999) or Pauker et al. (2011). That is, an early prosodic cue supports low attachment (pause after NP1) but later prosodic information (boundary tone on NP2) supported by syntactic information (Gender marking on the pronoun) disconfirms the initial analysis, thus leading to integration difficulty at this point. Unfortunately, the results are not discussed in these terms, probably because much of the ERP evidence about prosodic and syntactic processing (see 2.4.3.) wasn't available at the time.

Study	Language and task	Items and disambiguation	Effects
Augurzky (2005)	German Reading Listening	**1. Genitive** **a. synH proH[19]:** Das ist die Köchin des Wirts # **deren** Pudel nervtötend winselte. *This is the cook (fem; Nom) of-the publican (masc, Gen) whose (fem, Gen) poodle nervingly whimpered.* **b. synH proL:** Das ist die Köchin # des Wirts **deren** Pudel nervtötend winselte. **c. synL proH:** Das ist die Köchin des Wirts # **dessen** Pudel nervtötend winselte. **d. synL proL:** Das ist die Köchin # des Wirts **dessen** Pudel nervtötend winselte. *This is the cook (fem; Nom) of-the publican (masc, Dat) whose (fem, Gen) poodle nervingly whimpered*. **1. 'von'** **a. synH proH:** Das ist die Köchin von dem Wirt # **deren** Pudel nervtötend winselte. **b. synL proL:** Das ist die Köchin # von dem Wirt **deren** Pudel nervtötend winselte. **c. synH proH:** Das ist die Köchin von dem Wirt # **dessen** Pudel nervtötend winselte. **d. synL proL:** Das ist die Köchin # von dem Wirt **dessen** Pudel nervtötend winselte.	Reading Broadly distributed N400 for preferred HA independent of preposition Listening 1 CPS at NP2 (Wirt) for proH 2 CPS at NP1 (Köchin) and NP2 (Wirt) for proL Genitive: N400 for synHproL compared to synLproL 'von' N400 for synHproH compared to synLproH

[19] Syn = syntactic, pro = prosodic, H = high attachment, L = low attachment, # = pause , fem = feminine, masc = masculine, Nom = Nominative case, Gen = Genitive case

References

Augurzky, P. (2006). *Attaching relative clauses in German: The role of implicit and explicit prosody in sentence processing*. MPI for Human Cognitive and Brain Sciences.

Baggio, G., & Hagoort, P. (2011). The balance between memory and unification in semantics: A dynamic account of the N400. *Language and Cognitive Processes*, *26*(9), 1338-1367.

Ballinger, G. A. (2004). Using generalized estimating equations for longitudinal data analysis. *Organizational research methods*, *7*(2), 127-150.

Bieniek, M. M., Frei, L. S., & Rousselet, G. A. (2013). Early ERPs to faces: aging, luminance, and individual differences. *Frontiers in psychology*, *4*.

Bögels, S. (2011). The role of prosody in language comprehension: when prosodic breaks and pitch accents come into play. Unpublished PhD thesis.

Bögels, S., Schriefers, H., Vonk, W., Chwilla, D. J., & Kerkhofs, R. (2013). Processing consequences of superfluous and missing prosodic breaks in auditory sentence comprehension. *Neuropsychologia, 51,* 2715 – 2728.

Bögels, S., Schriefers, H., Vonk, W., Chwilla, D. J., & Kerkhofs, R. (2010). The interplay between prosody and syntax in sentence processing: The case of subject-and object-control verbs. *Journal of Cognitive Neuroscience*, *22*(5), 1036-1053.

Branigan, H. (2007). Syntactic priming. *Language and Linguistics Compass*, *1*(1-2), 1-16.

Breen, M., Watson, D. G., & Gibson, E. (2011). Intonational phrasing is constrained by meaning, not balance. *Language and Cognitive Processes*, *26*(10), 1532-1562.

Burns, R. P., & Burns, R. (2008). *Business research methods and statistics using SPSS*. Sage.

Carlson, K. (2009). How prosody influences sentence comprehension. *Language and Linguistics Compass*, *3*(5), 1188-1200.

Carlson, K., Clifton, C., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language*, *45*(1), 58-81.

Carlson, K., Clifton, C., & Frazier, L. (2009). Nonlocal effects of prosodic boundaries. *Memory & cognition*, *37*(7), 1014-1025.

Carreiras, M., & Clifton, C. (1993). Relative clause interpretation preferences in Spanish and English. *Language and Speech*, *36*(4), 353-372.

Carreiras, M., & Clifton, J. R. C.(1999). Another word on parsing relative clauses: Eyetracking evidence from Spanish and English. *Memory & Cognition*, *27*(5), 826-833.

Carreiras, M., Salillas, E., & Barber, H. (2004). Event-related potentials elicited during parsing of ambiguous relative clauses in Spanish. *Cognitive Brain Research*, *20*(1), 98-105.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological review*, *113*(2), 234.

Clifton Jr, C., & Frazier, L. (1988). Comprehending sentences with long-distance dependencies. In *Linguistic structure in language processing* (pp. 273-317). Springer Netherlands.

Clifton, C., Carlson, K., & Frazier, L. (2002). Informative prosodic boundaries. *Language and Speech*, *45*(2), 87-114.

Clifton, C., Carlson, K., & Frazier, L. (2006). Tracking the what and why of speakers' choices: Prosodic boundaries and the length of constituents. *Psychonomic bulletin & review*, *13*(5), 854-861.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic bulletin & review*, *12*(5), 769-786.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition*, *30*(1), 73-105.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21.

Desmet, T., De Baecke, C., Drieghe, D., Brysbaert, M., & Vonk, W. (2006). Relative clause attachment in Dutch: On-line comprehension corresponds to corpus frequencies when lexical variables are taken into account. *Language and Cognitive Processes*, *21*(4), 453-485.

Desmet, T., Brysbaert, M., & Baecke, C. D. (2002). The correspondence between sentence production and corpus frequencies in modifier attachment. *The Quarterly Journal of Experimental Psychology: Section A*, *55*(3), 879-896.

De Vincenzi, M., & Job, R. (1995). An investigation of late closure: The role of syntax, thematic structure, and pragmatics in initial interpretation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1303.

Eckstein, K., & Friederici, A. D. (2006). It's early: event-related potential evidence for initial interaction of syntax and prosody in speech comprehension. *Journal of Cognitive Neuroscience*, *18*(10), 1696-1711.

Eckstein, K., & Friederici, A. D. (2005). Late interaction of syntactic and prosodic processes in sentence comprehension as revealed by ERPs. *Cognitive Brain Research*, *25*(1), 130-143.

Farmer, T. A., Misyak, J. B., & Christiansen, M. H. (2012). Individual differences in sentence processing. *The Cambridge Handbook of Psycholinguistics*, 353.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of memory and Language*, *41*(4), 469-495.

Ferreira, F., & Clifton Jr, C. (1986). The independence of syntactic processing. *Journal of memory and language*, *25*(3), 348-368.

Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. In *Speech Prosody 2002, International Conference*.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 116-124.

Frazier, L. (1979). On comprehending sentences: Syntactic parsing strategies. *ETD Collection for University of Connecticut*, AAI7914150.

Frazier, L. (1987). Syntactic processing: evidence from Dutch. *Natural Language & Linguistic Theory*, *5*(4), 519-559.

Frazier, L., Carlson, K., & Clifton Jr, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, *10*(6), 244-249.

Frazier, L., & Clifton Jr, C. (1997). Construal: Overview, motivation, and some new evidence. *Journal of Psycholinguistic Research*, *26*(3), 277-295.

Frazier, L., Clifton Jr, C., & Carlson, K. (2004). Don't break, or do: prosodic boundary preferences. *Lingua*, *114*(1), 3-27.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive psychology*, *14*(2), 178-210.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in cognitive sciences*, *6*(2), 78-84.

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiological reviews*, *91*(4), 1357-1392.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1219.

Friederici, A. D., von Cramon, D. Y., & Kotz, S. A. (2007). Role of the corpus callosum in speech comprehension: interfacing syntax and prosody. *Neuron*, *53*(1), 135-145.

Gilboy, E., Sopena, J. M., Cliftrn Jr, C., & Frazier, L. (1995). Argument structure and association preferences in Spanish and English complex NPs. *Cognition*, *54*(2), 131-167.

Grassi, M., & Soranzo, A. (2009). MLP: A MATLAB toolbox for rapid and reliable auditory threshold estimation. *Behavior Research Methods*, *41*(1), 20-28.

Hagoort, P. (2003). How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *Neuroimage*, *20*, S18-S29.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in cognitive sciences*, *9*(9), 416-423.

Hagoort, P. (2007). The memory, unification, and control (MUC) model of language. *Automaticity and control in language processing*, 243-270.

Hardin, J. W., & Hilbe, J. M. (2003). *Generalized estimating equations*. John Wiley & Sons, Inc.

Hanley, J. A., Negassa, A., & Forrester, J. E. (2003). Statistical analysis of correlated data using generalized estimating equations: an orientation. *American journal of epidemiology*, *157*(4), 364-375.

Hemforth, B., Fernández, S., Clifton Jr, C., Frazier, L., & Konieczny, L. (2005). Relative clause attachment in German, English, Spanish and French: Effects of position and length. *Unpublished ms*.

Hemforth, B., Konieczny, L., & Scheepers, C. (2000). Syntactic attachment and anaphor resolution: The two sides of relative clause attachment. *Architectures and mechanisms for language processing*, 259-281.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, *30*(1), 47-61.

Holyk, G. G., & Pexman, P. M. (2004). The elusive nature of early phonological priming effects: Are there individual differences?. *Brain and Language*, *90*(1), 353-367.

House, D. (1990). *Tonal perception in speech*. Krieger Publishing Company.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language*, *59*(4), 434-446.

Jaeger, T. F., & Snider, N. (2008). Implicit learning and syntactic persistence: Surprisal and cumulativity. In *The 30th Annual Meeting of the Cognitive Science Society (CogSci08)* (p. 827).

Jun, S. A. (2003). Prosodic phrasing and attachment preferences. *Journal of Psycholinguistic Research*, *32*(2), 219-249.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., Mckeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(2), 163-178.

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*, *15*(1), 98-110.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and cognitive processes*, *15*(2), 159-201.

Kerkhofs, R., Vonk, W., Schriefers, H., & Chwilla, D. J. (2007). Discourse, syntax, and prosody: The brain reveals an immediate interaction. *Journal of Cognitive Neuroscience*, *19*(9), 1421-1434.

Kerkhofs, R., Vonk, W., Schriefers, H., & Chwilla, D. J. (2008). Sentence processing in the visual and auditory modality: Do comma and prosodic break have parallel functions?. *Brain research*, *1224*, 102-118.

Kidd, G. R., Watson, C. S., & Gygi, B. (2007). Individual differences in auditory abilities. *The Journal of the Acoustical Society of America*, *122*, 418.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*(2), 153-194.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, *59*, 1208.

Klatt, D. H., & Cooper, W. E. (1975). Perception of segment duration in sentence contexts. In *Structure and process in speech perception* (pp. 69-89). Springer Berlin Heidelberg.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23-49.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203-205.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in cognitive sciences*, *4*(12), 463-470.

Kutas, Marta, and Kara D. Federmeier. "Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP)." *Annual review of psychology* 62 (2011): 621-647.

Lee, E. K., & Watson, D. G. (2011). Effects of pitch accents in attachment ambiguity resolution. *Language and cognitive processes*, *26*(2), 262-297.

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *The Journal of the Acoustical Society of America*, *53*, 380.

Li, X., Hagoort, P., & Yang, Y. (2008). Event-related potential evidence on the influence of accentuation in spoken discourse comprehension in Chinese. *Journal of Cognitive Neuroscience*, *20*(5), 906-915.

Luck, S. J. (2005). An introduction to the event-related potential technique. Cambridge MA: The MIT Press.

Martin-Loeches, M., Hinojosa, J.A., Casado, P., Munoz, F., Fernandez-Frias, C., (2004). Electrophysiological evidence of an early effect of sentence context in reading. *Biological Psychology.* 65, 265–280.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*(3), 283-312.

Mitchell, D. C., Cuetos, F., Corley, M. M., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research*, *24*(6), 469-488.

Mo, Y., Cole, J., & Lee, E. K. (2008). Naive listeners prominence and boundary perception. *Proc. Speech Prosody, Campinas, Brazil*, 735-738.

Mueller, J. L., Hahne, A., Fujii, Y. and Friederici A. (2005). Native and nonnative speakers' processing of a miniature version of Japanese as revealed by ERPs. *Journal of Cognitive Neuroscience*. 17. 1229–44.

Murray, M. M., Brunet, D., & Michel, C. M. (2008). Topographic ERP analyses: a step-by-step tutorial review. *Brain topography*, *20*(4), 249-264.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of verbal learning and verbal behavior*, *22*(3), 358-374.

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 786.

Pannekamp, A., Toepel, U., Alter, K., Hahne, A., & Friederici, A. D. (2005). Prosody-driven sentence processing: An event-related brain potential study. *Journal of cognitive neuroscience*, *17*(3), 407-421.

Pauker, E., Itzhak, I., Baum, S. R., & Steinhauer, K. (2011). Effects of cooperating and conflicting prosody in spoken English garden path sentences: ERP evidence for the Boundary Deletion Hypothesis. *Journal of cognitive neuroscience*, *23*(10), 2731-2751.

Pernet, C. R., Chauveau, N., Gaspar, C., & Rousselet, G. A. (2011). LIMO EEG: a toolbox for hierarchical LInear MOdeling of ElectroEncephaloGraphic data. *Computational intelligence and neuroscience*, *2011*, 3.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. *Annals of the New York Academy of Sciences*, *999*(1), 58-75.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* (Doctoral dissertation, Massachusetts Institute of Technology).

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, *39*(4), 633-651.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: a critical review. *Psychological bulletin*, *134*(3), 427.

Pickering, M. J., & Van Gompel, R. P. (2006). Syntactic parsing. *Handbook of psycholinguistics*, *2*, 455-503.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America*, *90*, 2956.

Rousselet, G. A. (2012). Does Filtering Preclude Us from Studying ERP Time-Courses?. *Frontiers in Psychology*, *3*.

Schafer, A. J. (1997). *Prosodic parsing: The role of prosody in sentence comprehension* (Doctoral dissertation, University of Massachusetts Amherst).

Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, *29*(2), 169-182.

Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2005). Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. *Approaches to studying world situated language use: Psycholinguistic, linguistic and computational perspectives on bridging the product and action tradition*, 209-225.

Scheepers, C. (2003). Syntactic priming of relative clause attachments: Persistence of structural configuration in sentence production. *Cognition*, *89*(3), 179-205.

Sereno, S.C., Brewer, C.C., O'Donnell, P.J., (2003). Context effect in word recognition: evidence for early interactive processing. Psychological Science 14, 328–333.

Snedeker, J., & Casserly, E. (2010). Is it all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities. *Language and Cognitive Processes*, *25*(7-9), 1234-1264.

Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, *25*(2), 249-271.

Speer, S., Warren, P., & Schafer, A. (2003). Intonation and sentence processing. In *Proceedings of the Fifteenth International Congress of Phonetic Sciences, Barcelona, Spain*.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1521.

Spivey-Knowlton, M., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55*(3), 227-267.

Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. *The Oxford handbook of psycholinguistics*, *327*, 342.

Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and language*, *86*(1), 142-164.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature neuroscience*, *2*(2), 191-196.

Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and language*, *120*(2), 135-162.

Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, *136*(1), 64.

Traxler, M. J. (2007). Working memory contributions to relative clause attachment processing: A hierarchical linear modeling analysis. *Memory & cognition*, *35*(5), 1107-1121.

Tooley, K. M., & Traxler, M. J. (2010). Syntactic priming effects in comprehension: A critical review. *Language and Linguistics Compass*, *4*(10), 925-937.

Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, *39*(4), 558-592.

Truss, L. (2003). *Eats, Shoots & Leaves*, London: Profile Books.

Van Berkum, J. J., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of cognitive neuroscience*, *11*(6), 657-671.

Van Gompel, R. P., Pickering, M. J., Pearson, J., & Jacob, G. (2006). The activation of inappropriate analyses in garden-path sentences: Evidence from structural priming. *Journal of Memory and Language*, *55*(3), 335-362.

Van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, *45*(2), 225-258.

Van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. *Reading as a perceptual process*, 621-648.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and cognitive processes*, *25*(7-9), 905-945.

Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, *99*(2), B63-B72.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, *91*, 1707.