Overt prosody and plausibility as cues to relative-clause attachment in English spoken sentences

Daniela Zahn, Christoph Scheepers

We investigated the interplay between overt prosodic cues and semantic cues on the structural interpretation of spoken sentences that permit either high- or low-attachment of a final relative clause. Prosodic cues were manipulated via the presence or absence of a strong prosodic boundary before the relative clause, and semantic cues were induced via plausibility restrictions (e.g., *the servant of the actress who was* {*serving tea* / *very famous*}). In the first two experiments, each type of cue was studied in isolation while keeping influences of the relevant other cue constant. Experiment 1 employed a standard off-line comprehension task and suggested that prosodic cues were not as effective as semantic cues in biasing participants' attachment preferences. However, using a more implicit (and less biased) structural priming task, Experiment 2 showed that our overt prosody manipulation was actually no less effective than plausibility in biasing relative-clause attachments. Experiment 3 was, again, based on structural priming; here, the two factors were fully *crossed* to investigate the interaction between overt prosody and plausibility. This experiment showed that the two types of cues interact in a complex way, suggesting that (a) the amount of surprisal associated with cueing a generally dispreferred structure and (b) the type of revision necessary to resolve the ambiguity both play a major role in determining relative clause attachments.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1210v1 | CC-BY 4.0 Open Access | rec: 3 Jul 2015, publ: 3 Jul 2015

RUNNING HEAD: Overt prosody, plausibility, and relative-clause attachment

Overt prosody and plausibility as cues to relative-clause attachment in English spoken sentences

Daniela Zahn[1*], Christoph Scheepers[1]

[1]Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK,

*Correspondence: Daniela Zahn, Institute of Neuroscience and Psychology, University of Glasgow, 58 Hillhead Street, Glasgow, UK, G12 8QB
d.zahn.1@research.gla.ac.uk

Word count: 9562
Number of tables: 6
Number of figures: 4

Abstract

We investigated the interplay between overt prosodic cues and semantic cues on the structural interpretation of spoken sentences that permit either high- or low-attachment of a final relative clause. Prosodic cues were manipulated via the presence or absence of a strong prosodic boundary before the relative clause, and semantic cues were induced via plausibility restrictions (e.g., *the servant of the actress who was* {*serving tea* / *very famous*}). In the first two experiments, each type of cue was studied in isolation while keeping influences of the relevant other cue constant. Experiment 1 employed a standard off-line comprehension task and suggested that prosodic cues were not as effective as semantic cues in biasing participants' attachment preferences. However, using a more implicit (and less biased) structural priming task, Experiment 2 showed that our overt prosody manipulation was actually no less effective than plausibility in biasing relative-clause attachments. Experiment 3 was, again, based on structural priming; here, the two factors were fully *crossed* to investigate the interaction between overt prosody and plausibility. This experiment showed that the two types of cues interact in a complex way, suggesting that (a) the amount of surprisal associated with cueing a generally dispreferred structure and (b) the type of revision necessary to resolve the ambiguity both play a major role in determining relative clause attachments.

Keywords: *prosody*, *plausibility*, *relative-clause attachment*, *structural priming*, *ambiguity resolution*

1. Introduction

Language comprehension and production usually do not require much cognitive effort. This is remarkable given that language users have to integrate various linguistic (e.g. syntactic, semantic, pragmatic) and non-linguistic (e.g. world knowledge) constraints in a very short space of time. Hence, the question arises how this is done. Psycholinguists often address this question by focusing on specific information sources, and possible interactions between them, within a well-defined set of structures. Many of the investigated structures contain structural attachment ambiguities, which provide an excellent test case to study the relative importance of various constraints, and potential interactions between them. One frequently studied example of an attachment ambiguity is found in sentences comprising a complex noun phrase (NP) with an adjacent relative clause (RC), as in (1).

(1) *The criminal shot the servant of the actress who was almost deaf.*

The sentence is globally ambiguous because it remains unclear which part of the preceding complex object noun-phrase the relative clause refers to. If it modifies the entire noun-phrase, i.e. *the servant of the actress*, the relative clause attaches higher up in the syntactic tree (high-attachment, Figure 1), implying that *the servant* was almost deaf. By contrast, if the relative clause modifies the more recent noun phrase within the complex noun-phrase, i.e. *the actress*, it attaches lower down in the syntactic tree (low-attachment, Figure 2), implying that *the actress* was almost deaf.

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

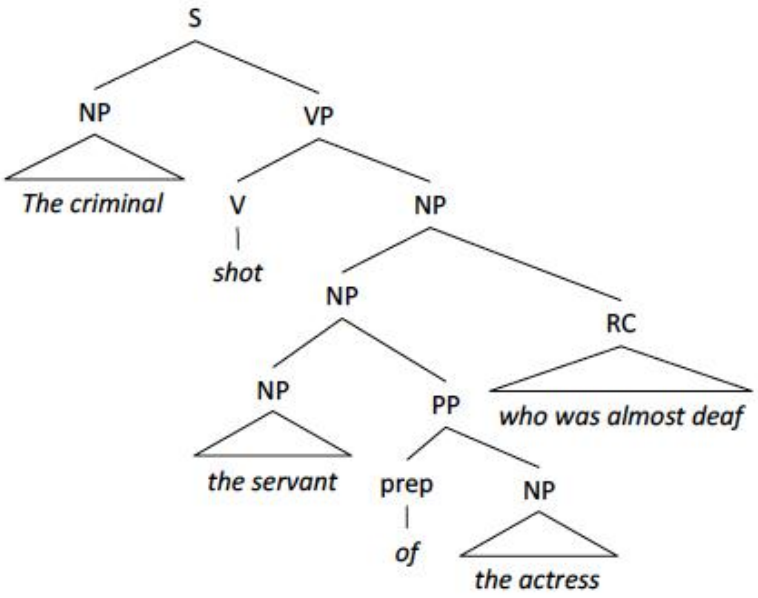
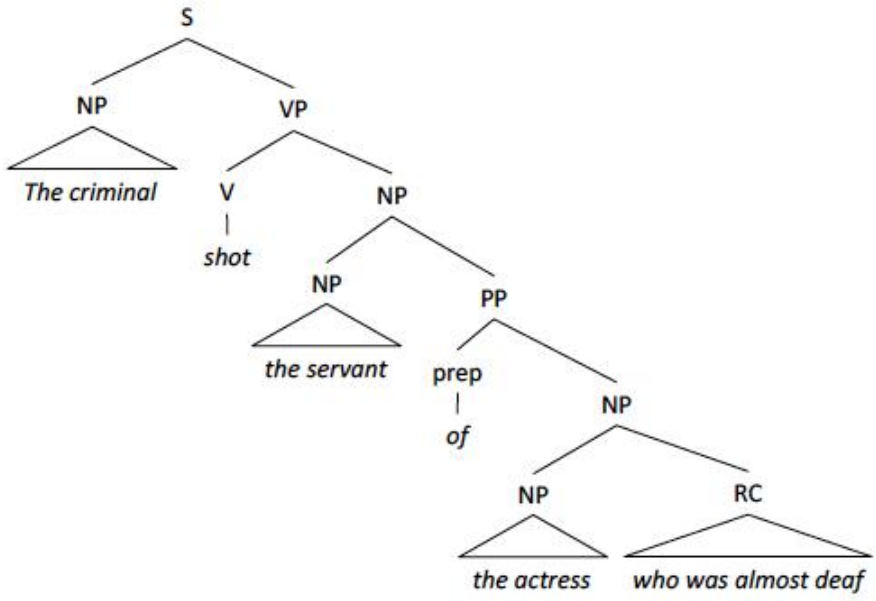


Figure 2. Phrase-structure representation for low-attachment (LA) of a relative clause (RC).



Readers or listeners are usually not aware of the ambiguity in sentences like (1) suggesting that such ambiguities are quickly resolved upon encounter. Native speakers of English tend to display a general low attachment preference for this kind of global syntactic ambiguity (e.g., Cuetos & Mitchell, 1988) which, in the first instance, would argue for universal processing heuristics such as *Late Closure* in Frazier's (1979, 1987a) Garden Path theory of sentence processing. 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 reported a high attachment preference for Spanish. If *Late Closure* were truly universal and apply to all languages, Spanish would display the same low attachment preference as English. Consequently, much research has focused on cross - linguistic differences in RC-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).

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 processing heuristics might offer a more suitable account for the data.

Despite the global syntactic attachment ambiguity in sentences such as 1, there are several ways of disambiguating them towards either one or the other interpretation of the relative clause.

Different aspects of semantic (e.g. animacy of the host noun phrases) and/or morpho-syntactic information 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 (e.g. Gilboy et al., 1995; Desmet et al., 2006). Take 2 (a, b) as an example:

2 *a. Someone shot the servant of the actress who was serving tea.*

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

Again, both sentences contain a complex noun phrase, i.e. *the servant of the actress*, which is modified by the following relative clause. The final two words within the relative clause in these examples provide semantic information making either high-attachment (2a) or low attachment (2b) more plausible (servants are more likely to serve tea and actresses more likely to be famous).

In research on reading, plausibility constraints such as in (2a,b) have frequently been used to maximally disambiguate relative-clause attachments in complex noun phrase structures (e.g. Carreiras & Clifton 1993; Cuetos & Mitchell, 1988; Gibson & Schuetze, 1999; Gilboy et al., 1995; Traxler, Pickering & Clifton, 1998; van Gompel, Pickering & Traxler, 2001; van Gompel et al., 2005) apparently because – in English at least – such constraints provide a very effective cue to the final high- or low-attachment interpretation of the relative clause.

In spoken language, overt prosodic features, such as the presence or absence of a prosodic boundary, have also been shown to bias the attachment of relative clauses in spoken sentences (Schafer et al. 1996; Clifton, Carlson, & Frazier, 2002). These overt prosodic features, and their interaction with plausibility constraints as in (2a,b), will be of primary interest in the present paper.[1]

Prosody refers to the rhythm, stress, and intonation of speech. Here, we will primarily focus on one particular aspect of overt prosody, which is roughly characterized 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. There are parameters that indicate a boundary such as a drop in F0 and an increase in pre-boundary syllable duration of the word before the boundary (House, 1990; Klatt, & Cooper,1975; Klatt,, 1976; Wightman et al., 1992). As such, they mark the boundaries of linguistically meaningful units as well as their prominence.

In spoken language, prosodic, syntactic and semantic cues are closely intertwined and notoriously difficult to disentangle when trying to understand their individual contributions as well as their interactions in establishing a coherent interpretation of a sentence. While studies into the prosody-syntax mapping show that this relationship is far from simple or conclusive (see Wagner & Watson, 2010 for a review), existing research points to a systematic relationship between overt prosody and syntax. Specifically, it has been shown that the prosodic structure of a spoken sentence has an influence on how listeners would parse such a sentence (e.g., Lehiste 1973; Beach, 1991; Cooper & Paccia-Cooper 1980; Price, Ostendorf, Shattuck-Hufnagel, Fong, 1991; Schafer, 1997; Carlson, Clifton & Frazier, 2001). Most relevant in the present context is a study by Clifton et al. (2002) who investigated, among other things, the influence of overt prosodic boundaries on the comprehension of relative-clause attachment ambiguities in English. Using an offline comprehension task, they found that an "informative" boundary (i.e. a boundary larger in size or prominence than any preceding prosodic boundary) before the relative clause in a spoken sentence such as (1) reliably biased listeners to assume a high-attachment interpretation of the relative clause (Clifton et al., 2002, Experiment 3).

However, to our knowledge, research in the auditory domain (e.g. Clifton et al. 2002; 2006) 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

[1] Note that we will not be concerned with *implicit* prosody in silent reading, although our investigations may have theoretical implications for the latter (see, e.g., Bader, 1998; Fodor 2002 a,b; Hirose, 2003; Traxler, 2009).

constraints were often used to maximally disambiguate relative-clause attachments while keeping (implicit) prosodic constraints more or less constant. Indeed, the question of how the two types of cues would cooperate in spoken language comprehension is interesting and important for at least two reasons: first, because it seems highly unlikely that the two types of constraints would always occur 'in isolation' in natural spoken language (i.e. outside a psycholinguistic laboratory); second, because an interaction between prosodic and semantic cues could reveal further insights into the relative salience of either type of cue in the interpretation of syntactically ambiguous spoken sentences.

In the first two experiments reported below, overt prosody and plausibility were manipulated in isolation (i.e., while keeping influences of the relevant 'other' cue constant) to evaluate their relative effectiveness in biasing globally ambiguous relative-clause attachments. Prosodic cues were manipulated via the presence or absence of a strong prosodic phrase boundary (i.e., a salient pause) before the critical relative clause, with the latter always being semantically neutral, as in (1). The theories discussed earlier all predict that the presence of such a pause before the relative clause (specifically when not preceded by another prominent prosodic boundary, as in our materials) should bias listeners to attach the ambiguous relative clause high, while the absence of such a boundary should support low-attachment of the relative clause. Semantic cues to relative-clause attachment were manipulated via pre-tested plausibility restrictions (cf. van Gompel, Pickering & Traxler, 2001; van Gompel et al., 2005) while keeping prosodic constraints constant (no prosodic boundary before the relative clause – note that it is difficult, if not impossible, to implement truly 'neutral' prosodic conditions in our materials). Experiment 1 employed a graded version of the two-alternative forced choice off-line comprehension task that is commonly used to establish attachment preferences for syntactically ambiguous sentences. Experiment 2 was based on a more implicit structural priming paradigm. As will be shown, type of task had a substantial impact on estimating the relative effectiveness of the two types of cues to relative-clause attachment, with structural priming being arguably less biased than standard off-line comprehension tasks.

Finally, Experiment 3 investigated structural priming of relative-clause attachments using a design in which overt prosody (pause present or absent) and plausibility (supporting high- or low-attachment of the relative clause) were *fully crossed* in the spoken prime materials. This way, we were able to examine potential interactions between the two types of constraints in biasing attachment preferences for syntactically ambiguous relative clauses, yielding a more ecologically valid assessment of the interaction of prosodic versus semantic cues in spoken language comprehension. Although the present research is to some extent exploratory, potential theoretical predictions will be discussed in the relevant experimental subsections.

2. Experiment 1

2.1 Design and Materials[2]

Twenty-four sets of materials were created. 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 (3) below. In the first two versions per item (3a and 3b), the relative clause was semantically 'neutral' in the sense that it could be a plausible modifier of both NP1 and NP2.

[2] Ethic approval has been obtained in 2010 from the Ethics committee of the Faculty of Information and Mathematical Science, Glasgow University (now Ethics Committee of the College of Science & Engineering). Unfortunately the reference number is no longer available.

The other two versions per item comprised relative clauses that would most plausibly combine with either NP1 (3c) or NP2 (3d). In other words, the final two versions semantically encouraged either high-attachment (3c) or low-attachment (3d) of the relative clause. In version (3a), the sentence was spoken such that there was a strong intonational phrase boundary (IPh)[3] before the critical relative pronoun, which is assumed to encourage high attachment of the relative clause. All other versions (3b-d) were spoken without a pause before the relative pronoun. This resulted in a two-factorial design comprising the factors *disambiguation* (prosodic [3a,b] versus semantic [3c,d]) and *attachment bias* (high [3a,c] versus low [3b,d]). A full list of experimental stimuli is provided in the Appendix, and example audio recordings are available at http://www.psy.gla.ac.uk/~danielaz/Audio_DZChS.zip.

(3) a. *The criminal shot the servant of the actress* [Pause] *who was almost deaf.*
b. *The criminal shot the servant of the actress who was almost deaf.*
c. *The criminal shot the servant of the actress who was serving tea.*
d. *The criminal shot the servant of the actress who was very famous.*

All items were recorded from a trained female native English speaker who was a graduate 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 strong intonational phrase boundary (marked by a falling intonation and a pause before the relative pronoun) in (3a) but no such boundary in (3b-c). We also ensured, via post-hoc editing of the audio files, that the pause before the relative pronoun in (3a) was held constant at 500 ms, and that no such pause occurred in the remaining conditions.

To confirm that the semantic and prosodic manipulations worked as intended, we carried out a plausibility rating study as well as acoustic analyses of the spoken materials.

2.1.1. Plausibility Pre-Test

We collected plausibility ratings to ensure that the relative clauses in condition (3c) semantically favored NP1 over NP2, that the relative clauses in condition (3d) semantically favored NP2 over NP1, and that the 'neutral' relative clauses in conditions (3a,b) combined equally well with both NP1 and NP2. 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 six different conditions for testing (i-vi).

(i) NP1-RC_{neut}: ... *a servant who is almost deaf.*
(ii) NP2-RC_{neut}: ... *an actress who is almost deaf.*
(iii) NP1-RC1: ... *a servant who is serving tea.*
(iv) NP2-RC2: ... *an actress who is very famous.*
(v) NP1-RC2: ... *a servant who is very famous.*
(vi) NP2-RC1: ... *an actress who is serving tea.*

The 24 (items) × 6 (conditions) = 144 stimuli were allotted into six 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

[3] For ease of terminology, we will henceforth refer to this boundary (and the following period of silence) as *Pause*.

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 six 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. As can be seen, the semantically neutral relative clauses in condition (i) and (ii) combined equally well with both NP1 and NP2. RC1 relative clauses (designed to semantically favor NP1) were significantly more plausible in combination with NP1 (iii) than with NP2 (vi). Conversely, RC2 relative clauses (designed to semantically favor NP2) were more plausible in combination with NP2 (iv) than with NP1 (v). Moreover, the semantically 'matching' conditions, (iii) and (iv), did not substantially differ from one another, and nor did the semantically 'mismatching' conditions, (v) and (vi). Overall, these results confirm that the semantic manipulations worked as intended.[4]

Table 1. Mean ratings (and standard errors) for each of the six NP-RC combinations used in the plausibility pre-test, together with results from *Tukey* tests (by subjects/items) comparing the conditions with one another.

		Tukey Tests by Subjects/Items						
Condition	Mean (SE)		(i)	(ii)	(iii)	(iv)	(v)	(vi)
NP1-RC_{neut}	4.47 (.08)	(i)	—	ns/ns	*/*	ns/ns	**/**	**/**
NP2-RC_{neut}	4.60 (.06)	(ii)		—	**/**	ns/ns	**/**	**/**
NP1-RC1	4.12 (.09)	(iii)			—	ns/ns	**/**	**/**
NP2-RC2	4.40 (.08)	(iv)				—	**/**	**/**
NP1-RC2	2.91 (.11)	(v)					—	ns/ns
NP2-RC1	2.90 (.10)	(vi)						—

**: $p < .001$; *: $p < .05$; ns: $p > .1$

2.1.2. Acoustic Analysis

Apart from inserting a pause before the relative pronoun in the high-attachment prosodic disambiguation condition (3a in *Design and Materials*), we examined acoustic parameters on the noun before the relative pronoun (N2) that are also commonly assumed to mark the presence of an IPh (House, 1990; Klatt, & Cooper, 1975; Klatt, 1976; Wightman et al., 1992). These include the duration of the stressed and last syllable of N2, as well as the fundamental frequency (F0) at the end of N2. The results from these analyses (carried out in Praat; Boersma, 2002) are summarised in Table 2.

[4] The plausibility ratings for the NP1-conditions (i) and (iii) were numerically lower than for the NP2-conditions (ii) and (iv). This could be because many of the NP1s, but none of the NP2s in our materials comprised relational nouns (e.g. *brother*) which prefer to occur in combination with prepositional phrases (e.g. *the brother of the girl*) rather than on their own (as in this pre-test) before being modified with a relative clause.

Table 2. Mean duration (in ms) of (a) the stressed and (b) the final syllable of N2, as well as (c) F0 in Hz at the end of N2, separately for each experimental condition. Also shown are the results from pair-wise *Tukey* tests by items. Condition labels (3a-d) correspond to the examples in the *Design and Materials* section.

			Tukey Tests by Items (*p*-values)				
	Condition	Mean (SE)		(3a)	(3b)	(3c)	(3d)
(a)	Pros-HA	270 (24)	(3a)	–	.94	.12	.99
	Pros-LA	255 (15)	(3b)		–	.34	.82
	Sem-HA	237 (15)	(3c)			–	.09
	Sem-LA	260 (16)	(3d)				–
	Condition	Mean (SE)		(3a)	(3b)	(3c)	(3d)
(b)	Pros-HA	356 (18)	(3a)	–	.001	.001	.001
	Pros-LA	259 (18)	(3b)		–	.99	.99
	Sem-HA	266 (16)	(3c)			–	.99
	Sem-LA	263 (18)	(3d)				–
	Condition	Mean (SE)		(3a)	(3b)	(3c)	(3d)
(c)	Pros-HA	146 (2)	(3a)	–	.001	.001	.001
	Pros-LA	175 (3)	(3b)		–	.005	.01
	Sem-HA	165 (2)	(3c)			–	.89
	Sem-LA	169 (2)	(3d)				–

As can be seen, there were no substantial differences in the duration of the stressed syllable of N2, but very clear differences in the duration of the last syllable of N2 (the former and the latter were identical only for items with monosyllabic N2s): The last syllable was longer in the high-attachment prosodic disambiguation condition (3a) than in the remaining conditions (3b-d). Correspondingly, F0 was significantly lower in the high-attachment prosodic

disambiguation condition (3a) than in the remaining conditions (3b-d). These parameters, as well as the pause before the relative pronoun, should all contribute to the perception of a strong prosodic boundary in (3a).

Finally, a trained phonetician (T.R., Glasgow University) analysed a random sample of 14 sound files for pitch accents and found that these comprised H* on NP1 and either H* or H+L* on NP2. Thus, differences in pitch accents are unlikely to account for any effects observed.

2.2. Procedure

The main part of Experiment 1 was conducted in DMDX (Forster & Forster, 2003). The 24 (items) × 4 (conditions) = 96 stimuli were placed into four presentation files using a Latin square (six items per condition per file). Also included were 52 structurally unrelated filler sentences recorded from the same speaker as the main items. Experimental sentences and fillers were presented in a pseudo-random order. Each presentation file started with three filler sentences. The experimental sentences were separated by at least two fillers. All sentences were presented acoustically via headphones while participants fixated a cross on the screen. Each experimental and filler sentence was then followed by a comprehension question, and two possible answers were given. In case of experimental sentences, the screen displayed, e.g., "*Who was almost deaf? The servant* << >> *The actress*". Participants had to indicate whether the relative clause modified NP1 or NP2 by providing a rating on a four-point scale. If participants were entirely certain that the relative clause modified NP1 (high-attachment), i.e. that *the servant was almost deaf*, they were asked to press 1 on the keyboard in front of them. If they were not entirely sure, but leaned towards high-attachment, they were asked to press 2. The same held for low-attachment of the relative clause, using the keys 3 (leaning towards low-attachment) and 4 (certain low-attachment), respectively. Note that this task is largely comparable to the two-alternative forced choice comprehension questions typically employed in this kind of research, except for the more 'graded' distinction between the two comprehension alternatives in our study.

2.3. Participants

Thirty English native speakers (21 females) participated in the main part of Experiment 1 in exchange for course credits. There were at least seven participants per presentation file. A typical session took about 10-15 minutes to complete.

2.4. Results

The mean ratings per condition (and corresponding standard errors) are shown in Table 3. As expected, the high-attachment bias (HA) conditions were associated with lower scores and the low-attachment bias (LA) conditions with higher scores. However, the difference was numerically smaller with prosodic than with semantic disambiguation.

Table 3. Mean attachment ratings (with standard errors) in each condition of Experiment 1. The critical sentences were either prosodically (via a pause or no pause) or semantically (via plausibility constraints) biased towards either high- or low-attachment of the RC. The scale ranged from 1 (certain high-attachment) to 4 (certain low-attachment).

Condition		Mean (SE)
Prosodic	HA-bias	2.34 (.09)
	LA-bias	2.68 (.09)

Semantic	HA-bias	1.54 (.07)
	LA-bias	3.51 (.07)

Statistical analyses were based on full-factorial mixed model ANOVAs. *Attachment-bias* (high-attachment vs. low-attachment) and *disambiguation* (prosodic vs. semantic) were entered as fixed factors, and either participant (F_1) or item (F_2) as random factors. In line with the descriptive pattern, the analyses revealed a significant main effect of *attachment-bias* ($F_1(1,29) = 247.15$; $p < .001$; $F_2(1,23) = 147.45$; $p < .001$) which was modulated by a reliable *disambiguation* × *attachment-bias* interaction ($F_1(1,29) = 120.38$; $p < .001$; $F_2(2,23) = 49.74$; $p < .001$).

Testing the simple effect of *attachment-bias* at each level of *disambiguation* showed that it was significant in both the semantic disambiguation condition ($F_1(1,29) = 717.52$; $p < .001$; $F_2(1,23) = 148.99$; $p < .001$) and the prosodic disambiguation condition ($F_1(1,29) = 6.43$; $p < .02$; $F_2(1,23) = 7.23$; $p < .02$). However, the significant two-way interaction indicates that the effect of semantic disambiguation was substantially larger than the effect of prosodic disambiguation.

2.5. Discussion

Both modes of disambiguation (i.e. overt prosody and plausibility) were found to be effective in biasing interpretation towards either high- or low attachment of the relative clause in a complex NP1-of-NP2-RC noun phrase. However, a reliable *disambiguation by attachment-bias* interaction suggested that plausibility is a much more effective disambiguation cue than overt prosody. Could this mean that the latter plays less an important role in the disambiguation of relative-clause attachments in spoken sentences?

A little reflection on the demands imposed by the task reveals that such a conclusion may be premature. First, memory-traces for the two types of linguistic information (prosodic vs. semantic) may differ in strength. For example, Mehler (1963) and Craik & Tulving (1975) found that the semantic content of a proposition is remembered better than its form. Second, and more importantly, the comprehension questions themselves may have biased the results. Recall that the relative clauses in the prosodic disambiguation conditions were always semantically unbiased. Hence, when answering correspondingly neutral comprehension questions in these conditions (e.g. "*Who was almost deaf? The servant << >> The actress*"), participants' interpretational biases could only have been influenced by the prosodic representations retained in memory from the previous auditory sentences. In stark contrast, the questions in the semantic disambiguation conditions effectively re-introduced the plausibility constraints from the previous sentences (e.g., "*Who was serving tea? The servant << >> The actress*"), so that participants could have answered them even without relying on what they had heard before. Taken together, the present experimental task is likely to have given plausibility constraints an advantage over prosodic constraints.

The following experiment employed a *structural priming* paradigm. Structural priming relies on the well-documented fact that language producers tend to repeat structures that they have uttered or encountered before (see Pickering & Ferreira, 2008). Indeed, there is clear evidence that relative-clause attachments are subject to this kind of priming as well, making this paradigm particularly useful for present purposes. Specifically, it was found that exposure to an unambiguous high-attached relative clause in a prime trial increases the probability of producing a high-attached relative clause in a subsequent target trial, and conversely, exposure to an unambiguous low-attached relative clause increases the likelihood of subsequently producing a low-attached relative clause (e.g. Scheepers, 2003; Desmet & Declercq, 2006).

In Experiment 2, the spoken sentence materials from Experiment 1 were used as *prime sentences*. These were followed by written target sentence fragments (to be completed verbally by the participants) which permitted both high- and low-attached relative clause continuations (e.g. "*The tourist guide mentioned the bells of the church that …*"). This task is arguably more implicit and less biased than the previous one because, for the disambiguation cues in the spoken sentences to be effective (as structural primes of subsequent target completions), participants can only rely on memory representations of those spoken sentences; there are no questions that could re-introduce any of the previous disambiguation cues.

3. Experiment 2

3.1. Participants

Thirty-two native English speakers (19 females) participated 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 end of each session.

3.2. Design and Materials

Experiment 2 was based on structural priming. The 96 spoken stimuli from Experiment 1 (24 items, four versions each) were used as primes. These were paired with 24 written sentence fragments like (4) which were used as targets. The same target was used across the four prime conditions per item (see Appendix).

(4) *The tourist guide mentioned the bells of the church that …*

The target fragments were unrelated in semantic content to the prime sentences. They always contained a complex NP1-of-NP2 noun phrase in object position, followed by a relative pronoun (*that* or *who*) and a "to-be-continued" marker (…) at the end. The critical host noun phrases in the target fragments (NP1 and NP2) were either both animate or both inanimate. They differed in number (NP1-plural versus NP2-singular in one half of the items and NP1-singular versus NP2-plural in the other half) which aided later classification of responses (see *Response Annotation*).

In addition to the prime-target pairs, 26 auditory sentences (recorded from the same speaker as the prime sentences) and 26 written sentence fragments were prepared as filler materials. Half of the spoken filler sentences were mildly implausible, mainly to motivate the task (see *Procedure*). The fillers were unrelated in content and syntactic structure from the experimental items so as to distract from the main purpose of the experiment.

3.3. 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, and 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.

Over 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 about 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.

3.4. Response Annotation

There were 766 useable target sound recordings (two trials were excluded due to incomplete recordings). The 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. As explained in *Design and Materials*, 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 $\kappa = .81$ (± .03 SE). This confirms the validity of the first annotator's classifications which were used in the main analyses.

3.5. Results

Overall, 181 (24%) of the valid target responses were classified as HA, 423 (55%) as LA, and 162 (21%) as UC. Hence, there was a general preference for low-attachment, consistent with earlier findings in English. Table 4 shows the target response distributions in each prime condition. As can be seen, there were proportionally more HA target responses after prime sentences that were prosodically or semantically biased towards high-attachment of the relative clause, and proportionally more LA target responses after prime sentences that were prosodically or semantically biased towards low-attachment of the relative clause.

Table 4. Probabilities of HA, LA, and UC target completions in each prime condition of Experiment 2 (raw counts in parentheses). The prime sentences were either prosodically (via a pause or no pause) or semantically (via plausibility constraints) biased towards either high- or low-attachment of the RC.

Prime Condition		Target Completion HA	LA	UC
Prosodic	HA-bias	.28 (53)	.52 (100)	.20 (39)
	LA-bias	.20 (38)	.63 (119)	.17 (33)
Semantic	HA-bias	.28 (54)	.50 (96)	.22 (42)
	LA-bias	.19 (36)	.56 (108)	.25 (48)

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 *disambiguation* (prosodic vs. semantic) and *attachment-bias* (high vs. low) were included as repeated-measures variables in a full-factorial 2 × 2 design using participants (χ^2_s), respectively items (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 – at most, there was a marginal *disambiguation* main effect by items ($\chi^2_s(1) = 2.15$; $p = .14$; $\chi^2_i(1) = 2.76$; $p = .097$). UC responses were therefore not considered further.

The next set of analyses focused on the proportions of HA target responses out of all *classifiable* target responses (HA and LA responses combined). This analysis showed a clear main effect of *attachment-bias* ($\chi^2_s(1) = 8.98$; $p < .005$; $\chi^2_i(1) = 7.47$; $p < .01$), but no effect of *disambiguation* (*p*s > .5) and no interaction between the two factors (*p*s > .7).

Ninety-five percent confidence intervals for the *attachment-bias* simple effect in each *disambiguation* condition (derived from the GEE model parameters) confirmed that prosodic cues (.10 ± .08 by subjects; .11 ± .07 by items) were no less effective than semantic cues (.11 ± .09 by subjects; .09 ± .09 by items) in priming subsequent target relative-clause attachments.

3.6. Discussion

Using a more implicit and less biased structural priming paradigm, Experiment 2 confirmed the effectiveness of overt prosody and plausibility as disambiguation cues for relative-clause attachments in English complex noun phrases. In contrast to the first experiment, Experiment 2 revealed no indication of either of these cues having a stronger impact than the other: primes whose relative-clause attachments were disambiguated via overt prosody were no less effective in biasing subsequent target RC-attachments than primes whose relative-clause attachments were disambiguated via plausibility.

Thus, with an unbiased task (structural priming), overt prosody and plausibility exert comparable relative-clause attachment biases when the influence of the relevant other cue is held constant. Now the question arises as to how the two types of cues would operate in a *fully crossed* experimental design in which they would either agree or disagree in their

support for high versus low attachment of the relative clause. This was examined in Experiment 3, using the same structural priming method as in Experiment 2. To our knowledge, crossing of prosody and plausibility cues to relative-clause attachment has not been studied or discussed before, making it difficult to generate specific theoretical predictions. The perhaps most parsimonious hypothesis would be that the two types of priming cues operate in an *additive* fashion so that high-attachment of the target-RC should be (i) most frequent when prosody and plausibility of the prime agree in their support for high-attachment, (ii) least frequent when prosody and plausibility of the prime agree in their support for low-attachment, and (iii) of intermediate frequency when prosody and plausibility of the prime support different attachments. This would predict two main effects, but no interaction between the two priming manipulations.

4. Experiment 3

4.1. Participants

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.

4.2. Design and Materials

Experiment 3 employed the same syntactic priming method as Experiment 2. The 96 spoken prime stimuli from Experiment 2 were cross-spliced to create four new priming conditions, as shown in (5a-d). The beginnings of the new prime stimuli (e.g. "*The criminal shot the servant of the actress* [Pause]" or "*The criminal shot the servant of the actress*") were taken from the previous prosodic prime items (3a) and (3b), respectively. The relative clauses (e.g. "*who was serving tea"* or "*who was very famous*") were taken from the previous semantic prime items (3c) and (3d), respectively. This resulted in a two-factorial design in which *prosody* (pause vs. no pause) and *plausibility* (high vs. low-attachment bias) were *crossed* to investigate potential interactions between the two types of cues.

(5) a. *The criminal shot the servant of the actress* [Pause] *who was serving tea.*
b. *The criminal shot the servant of the actress* [Pause] *who was very famous.*
c. *The criminal shot the servant of the actress who was serving tea.*
d. *The criminal shot the servant of the actress who was very famous.*

Note that all four conditions were created via cross-splicing. Thus, although conditions (5c) and (5d) were essentially the same as (3c) and (3d) in the previous experiments, the corresponding sound files were not identical. The main advantage of cross-splicing over new recordings is that acoustic parameters before and after the onset of the relative clause remain maximally comparable across conditions and experiments. The main disadvantage is that cross-splicing could introduce acoustic artifacts that might interfere with the priming effects of interest. The latter was addressed in an additional rating study.

4.3. Naturalness Ratings

To ensure that structural priming results could not be attributed to (or masked by) potential cross-splicing artifacts, we collected naturalness ratings from an additional 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 (example filler items can be downloaded at http://www.psy.gla.ac.uk/~danielaz/Audio_DZChS.zip). 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 judgments, participants had to press a corresponding number key (1-5) on the keyboard.

Overall, the critical items (5a-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* (*p*s < .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 (5a,b) were rated as slightly less natural/more likely to be edited (*M* = 2.78; *SD* = 1.63) than the two no-pause conditions (5c,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 *p*s > .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 main experiment were instructed to pay attention to whether the spoken sentences made sense or not (in line with the procedures in Experiment 2). Taken together, it seems unlikely that the results of the main experiment would be affected by cross-splicing artifacts in the primes.

4.4. Procedure and Response Annotation

The same target and filler materials, procedures, and response annotation criteria as in Experiment 2 were used. There were 40 participants × 24 items = 960 useable target sound recordings. Again, a random sample of 300 target completions (31%) was coded by a second annotator, yielding an inter-annotator agreement of 88% and $\kappa = .82$ (± .028 SE).

4.5. Results

In total, 292 (30%) of the valid target responses were classified as HA, 448 (47%) as LA, and 225 (23%) as UC. Table 5 shows the target response distributions in each prime condition.

Table 5. Probabilities of HA, LA, and UC target completions in each prime condition of Experiment 3 (raw counts in parentheses). The prime sentences contained either a pause or no pause before the RC and were semantically biased towards either high- or low-attachment of the RC (factorial design crossing prosody with plausibility).

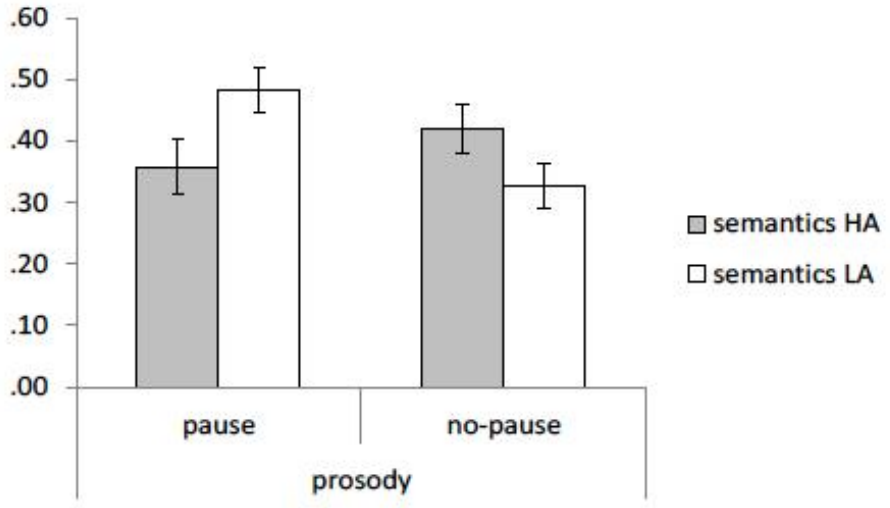
		Target Completion		
Prosody	Plausibility	HA	LA	UC
Pause	HA-bias	.28 (67)	.51 (122)	. 21 (51)
	LA-bias	.38 (91)	.40 (97)	. 22 (52)

No-Pause	HA-bias	.31 (74)	.44 (106)	. 25 (60)
	LA-bias	.25 (60)	.51 (123)	.24 (57)

As before, binary logistic GEE modeling was employed. Prime *prosody* (pause vs. no-pause) and *plausibility* (high- vs. low-attachment bias) were included as within-subjects (χ^2_s) respectively within-items (χ^2_i) predictors in a full-factorial 2 × 2 design, assuming an exchangeable covariance structure for repeated measurements. The *Generalized Score Chi Square* statistic was used for hypothesis testing.

Analyses of unclassifiable (UC) target responses in proportion to all available responses established no significant effects (all *p*s > .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 *p*s > .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 (5c vs. 5d) replicated the plausibility-driven priming effect from Experiment 2, showing more HA target completions when the prime-RC was semantically biased towards high- (5c) than towards low-attachment (5d); 95% CIs for the simple effect: .09 ± .08 by subjects; .08 ± .08 by items. Intriguingly, the comparison between the two pause conditions (5a vs. 5b) showed a *reverse* simple effect of *plausibility*, with reliably fewer HA target completions when the prime-RC was semantically biased towards high- (5a) than towards low-attachment (5b); 95% CIs: −.13 ± .10 by subjects; −.11 ± .10 by items. Figure 3 plots the estimated marginal means (with by-subject *SE*s) per condition.

Figure 3. Proportions of HA target completions (out of all classifiable completions) as a function of prime *prosody* and *plausibility* in Experiment 3. Error bars represent by-subject standard errors for the means.



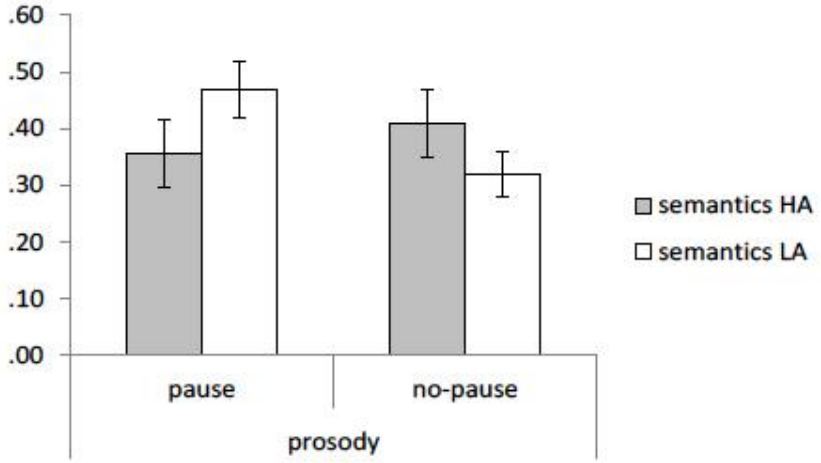
Although the naturalness ratings did not suggest very strong cross-condition differences, we performed a supplementary analysis to establish whether the above structural priming results were in any way influenced by the perceived naturalness of the primes. To this end, the naturalness ratings were aggregated into item-by-condition means and used as an additional covariate in a binary logistic GEE analysis on proportions of HA out of all classifiable (HA

and LA) target completions. Since the ratings were from a different participant sample than the priming data, this analysis was by-items only. The corresponding inferential results are summarised in Table 6; Figure 4 shows the relevant covariate-adjusted means with by-item *SE*s.

Table 6. Inferential results from the supplementary binary logistic GEE analysis (by items only) on proportions of HA target responses in Experiment 3. The analysis was based on a full-factorial *naturalness* (covariate) × *plausibility* (HA-bias vs. LA-bias) × *prosody* (pause vs. no-pause) design, assuming an exchangeable covariance structure for repeated measurements; χ^2 refers to the *Generalized Score Chi-Square* statistic for hypothesis testing.

GEE Effect Term	$\chi^2(1)$	*p*
naturalness	0.40	.53
plausibility	0.12	.73
prosody	2.42	.12
naturalness × *plausibility*	0.08	.77
naturalness × *prosody*	0.92	.34
prosody × *plausibility*	6.57*	.01
naturalness × *prosody* × *plausibility*	0.16	.69

Figure 4. Covariate-adjusted proportions of HA target completions (out of all classifiable completions) as a function of prime *prosody* and *plausibility* in Experiment 3 (supplementary item analysis). Error bars represent by-item standard errors for the means.



As can be seen, inclusion of the *naturalness* covariate did not diminish the significance of the *prosody* × *plausibility* interaction. The covariate itself did not explain much variability in the structural priming data, as evidenced by the lack of any significant effect terms involving *naturalness*.

4.6. Discussion

In the first two experiments, we looked at the effectiveness of prosodic and semantic cues to relative-clause attachment 'in isolation', that is, by keeping the impact of the relevant other cue constant. The present structural priming study (Experiment 3) investigated their combined influences in a fully *crossed* experimental design using cross-spliced materials from the previous experiments as primes.

Experiment 2 had shown that overt prosody and plausibility were equally effective in priming subsequent target-RC attachments. Therefore, the most parsimonious prediction for the combined effect of the two types of priming cues would be that they operate in an *additive* fashion, producing the strongest priming effects whenever they agree in their support for a given attachment.

This prediction was clearly not confirmed. Instead, Experiment 3 revealed a rather interesting interaction between the two types of cues in the prime: without a pause before the prime-RC (prosodic support for low-attachment), semantic cues worked in the expected direction, consistent with the plausibility-driven priming effect in Experiment 2; however, when the prime-RC was preceded by a pause (prosodic support for high-attachment), the effect of plausibility was *reversed*, showing stronger high-attachment priming when plausibility favored low-attachment of the relative clause. Importantly, this pattern of results is unlikely to be due to cross-splicing artifacts: first, the two no-pause conditions (5c vs. 5d, based on cross-spliced materials) replicated the plausibility-driven priming effect observed in Experiment 2 (3c vs. 3d, based on natural recordings); second, the accompanying rating study suggested only small and rather inconsistent differences in perceived naturalness across the four priming conditions; finally, using the naturalness ratings as an additional covariate in the by-item analysis showed no appreciable relationship between the covariate and the priming data. In all likelihood, the observed *prosody* × *plausibility* interaction therefore reflects genuine non-additivity in the combination of overt prosodic and semantic constraints on relative-clause attachment, which deserves thorough consideration in the general discussion.

5. General Discussion

Over thee experiments, we investigated two different modes of disambiguating relative clause-attachments within spoken NP1-of-NP2-RC (e.g. "the servant of the actress who …") noun phrases – both in isolation (Experiment 1 and 2) and in a *fully crossed* experimental design (Experiment 3). The first mode of disambiguation was *overt prosody*, manipulated via the presence or absence of a pause before the relative clause. In line with earlier research (e.g. Clifton et al., 2002), we expected that the pause condition would bias listeners towards a high-attachment interpretation of the relative clause, whereas the no-pause condition should support the generally preferred (in English) low-attachment interpretation of the relative clause. Experiments 1 and 2 confirmed these predictions under conditions with equal semantic support for either type of attachment. The second mode of disambiguation was *plausibility*, as manipulated via (pre-tested) semantic restrictions between the relative clause and the preceding host noun phrases – a manipulation that is often used to 'maximally disambiguate' relative-clause attachments in reading research (e.g. Carreiras & Clifton 1993; Cuetos & Mitchell, 1988; Gibson & Schuetze, 1999; Gilboy et al., 1995; Traxler, Pickering & Clifton, 1998; van Gompel, Pickering & Traxler, 2001; van Gompel et al., 2005). Indeed, Experiments 1 and 2 confirmed the effectiveness of this manipulation also for spoken sentence processing, and under conditions where prosodic constraints on relative-clause attachment were held constant (no pause before the relative clause).

While both modes of disambiguation consistently resulted in the expected relative-clause attachment preferences, Experiment 1 suggested that prosodic cues influence such biases to a lesser extent than semantic cues, as indicated by a significant *disambiguation* ×

attachment bias interaction. However, this could largely be attributed to the experimental task (two-alternative forced choice question-answering) which gave semantic constraints an advantage over prosodic constraints by effectively re-introducing the plausibility constraints in the question itself. Interestingly, recent research on the role of pitch accent on ambiguity resolution points to similar artifacts induced by this kind of task (Lee & Watson, 2011). In Experiment 2, the spoken materials from the first experiment were used as primes in a more implicit (and arguably less biased) structural priming task. This experiment showed that in our materials, overt prosody was actually no less effective than plausibility in biasing participants' preferred relative-clause attachments: the two modes of disambiguation in the spoken prime sentences lead to comparable priming effects in the subsequent target trials. Given its implicit and unbiased nature, structural priming therefore appears to be an extremely useful addition to the inventory of methods that probe into the relative effectiveness of different modes of syntactic disambiguation.

Experiment 3 employed the same structural priming paradigm to address the theoretically most interesting question of this paper, namely 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. Given that the two types of cues were found to be equally effective primes of subsequent target RC-attachments when studied 'in isolation' (Experiment 2), a parsimonious prediction might 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). Clearly, this prediction turned out to be too simplistic. Instead of two main effects, we found a rather interesting interaction between the two types of cues in Experiment 3: Without a pause before the relative clause, plausibility restrictions in the spoken prime sentence biased target relative-clause attachments in the expected manner (more high-attached target-RCs when plausibility restrictions in the prime supported high-attachment of the relative clause, consistent with the findings from Experiment 2); however, when there was a pause before the relative clause in the prime, the effect of plausibility was *reversed*, yielding more high-attached target-RCs when plausibility restrictions in the prime supported *low-attachment* of the relative clause. Given that cross-splicing artifacts were unlikely to be an issue (see Discussion of Experiment 3), one explanation of this interaction might rely on the notion of *surprisal* associated with a given disambiguation cue (cf. Jaeger & Snider, 2008; Scheepers, 2003). Assuming that low-attachment is generally preferred for the structures under investigation,[5] a cue in support of the alternative high-attachment interpretation should be more surprising – and thus more salient and effective in biasing target attachment decisions – than a cue that is in line with the general low attachment preference. This might explain why plausibility cues to high attachment seem particularly effective when overt prosodic cues support low attachment of the relative clause (no-pause conditions), and conversely, why overt prosodic cues to high attachment (pause before the relative clause) seem most effective when plausibility constraints are in line with the general low-attachment preference (which also includes semantically 'neutral' conditions, cf. Experiment 2).

[5] Indeed, a general low-attachment preference is not only suggested by prior research on relative-clause attachment in English (see Introduction), but also reflected in the fact that low-attachment (LA) responses accounted for more than 60% of all classifiable target responses in Experiments 2 and 3.

However, this kind of interpretation does not convincingly address one rather curious aspect of the observed *prosody* × *plausibility* interaction in Experiment 3 (cf. Figures 3 and 4): when *both* types of cues in the prime supported the less preferred high attachment interpretation of the relative clause (*pause, semantics HA* condition) then 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 (*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 cue alone (in terms of whether it disagrees with the general low-attachment preference) may not be sufficient to explain our data.

Instead, the results suggest an additional mediating factor which relates to (a) the relative point in time at which each type of cue becomes available in the spoken prime sentence and (b) the likelihood and type of a structural revision that has to take place in case of a clash between the two types of cues in the prime. With respect to timing, overt prosodic cues (presence or absence of a pause before the relative clause) are available earlier in the sound stream than plausibility cues (the latter are instantiated within the relative clause itself). Prosodic cues are therefore likely to determine *early* attachment decisions during auditory processing of the prime sentence, whereas plausibility constraints are considered later and may trigger a revision of those earlier parses suggested by prosody. Indeed, the two conditions in which the two types of cues in the prime support the same relative-clause attachments (i.e., *pause, semantics HA* and *no-pause, semantics 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. However, structural revisions are very likely in the remaining two conditions where prosody and plausibility support different RC-attachments (i.e., *pause, semantics LA* and *no-pause, semantics HA*). Note that both of these cue-conflict conditions elicited reliable high-attachment priming effects compared to the two no-conflict conditions – curiously, even when 'late' plausibility cues supported low-attachment of the relative clause (*pause, semantics LA* condition). This might be taken as an indication of two qualitatively different structural revision processes: in case of early prosodic support for high-attachment (*pause*) followed by late semantic support for low-attachment (*semantics LA*), the more informative/surprising prosodic cue ultimately 'wins', as reflected in more high-attachment target responses in that condition; in contrast, if early prosodic support for low-attachment (*no-pause*) is followed by late semantic support for high-attachment (*semantics HA*), the latter cue is more surprising, and therefore more decisive in priming subsequent target RC-attachments. In sum, the observed *prosody* × *plausibility* interaction in Experiment 3 seems to rely on a combination of (a) surprisal associated with a given disambiguation cue and (b) structural revision during auditory processing of the prime. The former determines the *likelihood* while latter predicts the *direction* of structural priming.

If this interpretation is correct, then it follows that late plausibility constraints are able to overrule the *absence*, but not the *presence* of a strong prosodic boundary before the relative clause in the spoken prime. Indeed, such a conclusion is not without precedent in the literature. Strong effects of pauses, especially in cases where the pause indicates a different syntactic configuration than subsequent information, have been reported earlier (e.g., Speer et al., 1996; Kjelgaard & Speer, 1999). Most interestingly, Pauker et al. (2011) recently reported an ERP study in which they manipulated the positioning of pauses within early versus late closure sentences (6):

(6) a. *When a bear is approaching* [Pause] *the people come running.*
b. *When a bear is approaching the people come running.*

c. *When a bear is approaching the people* [Pause] *the dogs come running.*
d. *When a bear is approaching* [Pause] *the people* [Pause] *the dogs come running.*

They found that the 'missing pause' in an early closure sentence (6b versus 6a) elicited a rather weak P600 response, suggesting relatively mild structural revision processes. In contrast, the second pause in late closure sentences (6d versus 6c) elicited a strong biphasic N400/P600 response, suggesting far more extensive reanalysis and repair processes. Pauker et al. (2011) offered an explanation that accounts not only for their own results but also for a number of previously reported ones (e.g., Speer et al., 1996; Kjelgaard & Speer, 1999; Walker et al., 2001). In their *Boundary Deletion Hypothesis*, they claim that the deletion of a misplaced intonational boundary (as in 6d) is harder to achieve than the retrospective mental insertion of a pause that was actually missing in the sound stream (as in 6b). They propose that *"... any attempt to mentally undo the 'positive evidence' of a boundary in the speech signal implies the listener's willingness to assume that the speaker mistakenly produced the salient boundary cues (compared to the more likely case of having missed an insufficient boundary marking...)*" (Pauker et al., 2011, p. 2748). The Boundary Deletion Hypothesis could explain why plausibility constraints are able to override the absence, but not the presence of a strong prosodic boundary cue in our relative-clause attachment primes in Experiment 3. Indeed, it would be interesting to investigate corresponding spoken materials using an on-line method such as EEG, as this might reveal more direct clues to the hypothesized structural revision processes than the structural priming paradigm is able to offer.

6. Conclusion

In this paper, we investigated the influence of overt prosodic cues and plausibility cues on the interpretation of spoken sentences that permit either high- or low-attachment of a final relative clause. The two types of cues were studied both 'in isolation' and in a fully crossed experimental design in which they either agreed or disagreed in supporting different relative-clause attachments. The latter is interesting because the two types of cues are unlikely to occur independently of one another in natural speech. While our findings clearly raise a number of interesting questions for future research, there were some important methodological and theoretical lessons to be learnt from the present investigations. One is that, in order to evaluate each individual disambiguation cue's contribution to relative-clause attachment, it is imperative to use a task that does not favor one cue over the other. Structural priming appears to be very useful for this purpose due to its implicit and unbiased nature. The second important conclusion is that overt prosody and plausibility interact in non-trivial ways in determining relative-clause attachment preferences for spoken sentences: while plausibility constraints in support of a (non-default) high-attachment interpretation are able to override the *absence* of a prosodic boundary (with the latter supporting low-attachment), the *presence* of a strong prosodic boundary before the relative clause (supporting high-attachment) cannot be overridden by plausibility. This can be viewed as an additional confirmation of Pauker et al.'s (2011) *Boundary Deletion Hypothesis*. The third noteworthy suggestion from our data is that structural priming is most likely to occur when the two types of cues in the prime *disagree* in their support for high or low attachment of the relative clause, pointing to the potential importance of structural revision in explaining the priming effects in Experiment 3. In conclusion, the present investigations mark a promising initial step towards understanding the interplay between overt prosody and plausibility as cues to relative-clause attachment in English spoken sentences.

References

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In J.D. Fodor & F. Ferreira (Eds.), *Reanalysis in Sentence Processing* (pp. 1-46). Dordrecht: Kluwer.

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language*, 30, 644-663.

Boersma, P. P. G. (2002). Praat, a system for doing phonetics by computer. *Glot International*. 5(9/10). 341-345.

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

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

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

Clifton, C. Jr., Carlson, K., & Frazier, L. (2002), Informative Prosodic Boundaries. *Language and Speech, 45(2)*, 87-114.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.

Craik, F. I. M. & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104 (3), 268-294.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: restrictions on the issue of the late closure strategy in Spanish. *Cognition*, 30, 73-105.

Desmet, T., Brysbaert, M., & De Baecke, C. (2002). The correspondence between sentence production and corpus frequencies in modifier attachment. *Quarterly Journal of Experimental Psychology*, 55A(3), 879-896.

Desmet, T., & Declercq, M. (2006). Cross-linguistic priming of syntactic hierarchical configuration information. *Journal of Memory and Language*, 54(4), 610-632.

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*, 453-485.

Fodor, J.D. (2002a). Psycholinguistics cannot escape prosody. Paper presented at Speech Prosody 2002, Aix-en-Provence, France, April 11-13.

Fodor, J.D. (2002b). Prosodic disambiguation in silent reading. In: *Proceedings of the North East Linguistic Society*, 32 (pp. 113-132). Amherst: GSLA, University of Massachusetts.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, 35*, 116–124.

Frazier L. (1987). Sentence processing. In: M. Coltheart (Ed.), Attention and performance XII: The psychology of reading, Erlbaum, Hillsdale, N.J (1987)

Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.

Frazier, L., Carlson, K. & Clifton, C. (2006). Prosody is central to language comprehension. *Trends in Cogntive Science*, 10, 244-249.

Frazier, L., & Clifton, C. (1998). Sentence reanalysis and visibility. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 143-176). Dordrecht, the Netherlands: Kluwer.

Gibson, E. & Schuetze, C. T. (1999). Disambiguation preferences in noun phrase conjunction do not mirror corpus frequency. *Journal of Memory and Language, 40*, 263-279.

Gilboy, E., Sopena, J. M., Clifton, C. & Frazier, L. (1995). Argument structure and association preference in Spanish and English complex NPs. *Cognition*, *54*, 131-167.

Hanley, J. A., Negassa, A., Edwardes, M. D., & Forrester, J. E. (2003). Statistical analysis of correlated data using generalized estimating equations: An orientation. *American Journal of Epidemiology*, 157(4), 364-375.

Hardin, J. & Hilbe, J. (2003). *Generalized Estimating Equations*. London: Chapman and Hall/CRC.

Hemforth, B., Konieczny, L., Scheepers, C. & Strube, G. (1998). Syntactic ambiguity resolution in German. In D. Hillert (Ed.), *Sentence processing: A crosslinguistic perspective— syntax and semantics*, 31. San Diego, CA: Academic Press. 293-312.

Hemforth, B., Konieczny, L., & Scheepers, C. (2000). Syntactic attachment and anaphor resolution: the two sides of relative clause attachment. In M. Crocker, M. Pickering, & C. Clifton, Jr. (Eds.), *Architectures and mechanisms for language processing* (pp. 259-281). Cambridge: Cambridge University Press.

Hirose, Y. (2003). Recycling prosodic boundaries. *Journal of Psycholinguistic Research*, 32 (2), 167-195.

House, D. (1990). *Tonal Perception in Speech.* Lund: Lund University Press.

Jaeger, F. T. (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, F. T., & Snider, N. (2008). Implicit learning and syntactic persistence: Surprisal and Cumulativity. *The 30th Annual Meeting of the Cognitive Science Society* (CogSci08), 1061-1066.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic ambiguity. Journal of Memory and Language, 40, 153–194.

Klatt, D. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, *3*, 129-140.

Klatt, D. & Cooper, W. E. (1975). Perception of segment duration in sentence contexts. In: A. Cohen and S. Nooteboom (Eds.), *Structure and Process in Speech Perception* (pp. 69-89), Heidelberg: Springer.

Klatt, D. (1976): Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *JASA 59*, 1208-1221.

Lee, E., & 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. *Glossa*, 7, 107 - 122.

Marcus, M., & Hindle, D. (1990). Description theory and intonation boundaries. In G. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 483-512). Cambridge, MA: MIT Press.

Mehler J. (1963). Some effects of grammatical transformations on the recall of English sentences. *Journal of Verbal Learning and Verbal Behavior*, 2(4). 346-351.

Pauker, E., Itzhak I., Baum, S. R., & Steinhauer, K. (2011). Co-operating and conflicting prosody in spoken English garden path sentences: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, 23(10), 2731-2751.

Pickering, M. & Ferreira, V. (2008). Structural priming. A critical review. *Psychological Bulletin*, 134(3), 427-459.

Pierrehumbert, J. (1980*). The phonology and phonetics of English intonation.* PhD thesis. MIT. Indiana University Linguistics Club.

Price, P. J., Ostendorf, S., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, 9, 2956-2970.

Pynte, J., & Prieur, B. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes*, 11(1), 165-192.

Potter, M. C. & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, 29(6). 633-654.

Schafer, A. J., Carter, J., Clifton, C. & Frazier, L. (1996). Focus in relative clause construal. *Language and Cognitive Processes*, 11, 135-163.

Schafer, A. (1997). Prosodic parsing: The role of prosody in sentence comprehension. Doctoral dissertation, University of Massachusetts, Amherst.

Schafer, A. J., Speer, S. R., & Warren, P. (2005). Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. In M. Tanenhaus and J. Trueswell (Eds.), *Approaches to studying world situated language use* (pp. 209-225). Cambridge: MIT Press.

Scheepers, C. (2003). Syntactic priming of relative clause attachments: Persistence of structural configuration in sentence production. *Cognition*, 89, 179-205.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, 48, 103-130.

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*, 247-268.

Traxler, MJ. (2009). A Hierarchical Linear Modeling Analysis of Working Memory and Implicit Prosody in the Resolution of Adjunct Attachment Ambiguity. *Journal of Psycholinguistic Research*, 38, 491–509.

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, 558-592.

Van Gompel, R. P. G., Pickering, M. J., Pearson, J. & Liversedge, S.P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, 52, 284-307.

Van Gompel, R. P. G., 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, 225-258.

Walker, J., Fongemie, K., & Daigle, T. (2001). Prosodic facilitation in the resolution of syntactic ambiguities in subjects with left and right hemisphere damage. Brain and Language, 78, 169–196.

Watson, D., & Gibson, E. A. (2005). Intonational phrasing and constituency in language production and comprehension. *Studia Linguistica*, 59(2-3), 279-300.

Wagner, M. & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, 25(7&9), 905-954.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M. & Price, P. J. (1992). Segmental duration in the vicinity of prosodic phrase boundaries. *JASA 91*, 1707-1717.

Appendix

Transcripts of the spoken stimuli for Experiments 1 and 2

a. semantically neutral relative clause (used for prosodic disambiguation)
b. relative clause semantically biased towards high-attachment (HA)
c. relative clause semantically biased towards low-attachment (LA)

1. The criminal shot the servant of the actress who { a. was almost deaf., b. was serving tea., c. was very famous. }
2. The student thought about the content of the book that { a. interested him a lot., b. was very abstract., c. was rather heavy. }

3. Today Jane spoke with the father of the pupil who { a. was rude to the teacher., b. owned a shop., c. was doing well in class. }

4. The mechanic repaired the engine of the race car that { a. was developed very recently., b. had titanium pistons., c. had a new kind of spoiler. }

5. The politician referred to the source of the information that { a. was not reliable., b. had contacted him., c. was not newsworthy. }

6. Mary babysits the child of the musician that { a. was in the other room., b. was in the cot next-door., c. had a beard. }

7. I know the father of the secretary who { a. has a good sense of humour., b. is a retired policeman., c. who married a doctor. }

8. John detests the wife of the artist who { a. is ginger., b. is pregnant., c. wore a moustache. }

9. Someone smashed the window of the car that { a. was already damaged., b. was made of tinted glass., c. had a big exhaust. }

10. Paddy showed the costumer the mother of the puppy that { a. was brown., b. was old., c. was newborn. }

11. Peter approached the manager of the pop star who { a. was smoking a cigarette., b. formulated the contract., c. released a new album. }

12. The analyst commented on the development of the market that { a. was promising., b. was surprising., c. was growing. }

13. Eileen liked the colour of the dress that { a. was very much in fashion., b. was bright and fresh., c. was made of silk. }

14. Daniela was very happy about the funding of the project that { a. was approved last week., b. will be sufficient to pay the subjects., c. will be conducted within the department. }

15. The fans admired the coach of the wrestler who { a. retired after a long career., b. trained him for years., c. injured his knee. }

16. The board discussed the summary of the survey that { a. had been reported in the news., b. highlighted the most important points., c. was undertaken in the previous year. }

17. The scientist was pleased with the result of the experiment that { a. was novel and interesting., b. clearly confirm her prediction., c. uses a new kind of method. }

18. John argued with the brother of the girl who { a. was standing next to him., b. was a sales manager., c. was a bully at school. }

19. James knocked on the door of the house that { a. was painted red., b. was left ajar., c. had a new roof. }

20. The salesperson ignored the mother of the baby who { a. was looking at him., b. was searching her handbag., c. was making tantrums. }

21. The advisor commented on the progress of the work that { a. the company envisaged., b. appeared slower then expected., c. had been carried out recently. }

22. The PR manager looked at the advert of the company that { a. sold millions of IT products., b. appeared on the screen in front., c. owned a big production studio. }

23. The veterinarian examined the leg of the horse that { a. had sustained a severe injury., b. appeared to be broken., c. was supposed to win the race. }

24. The journalist interviewed the agent of the movie star who { a. had a strong accent., b. had made the contract., c. had won an Oscar. }

Target sentence fragments for Experiments 2 and 3

1. The tourist guide mentioned the bells of the church that ____ .
2. The manager waited for the musicians of the pop star who ____ .
3. The commission referred to the source of the donations that ____ .
4. Nora visited the students of the piano teacher who ____ .
5. The chauffeur met the representative of the state guests who ____ .
6. The tutor advised the students of the lecturer who ____ .
7. The superintendent checked the earnings of the company that ____ .
8. The bus driver talked to the leader of the boy scouts who ____ .
9. The farmhand fed the calves of the cow that ____ .
10. The pensioner complained about the content of the fliers that ____ .
11. The frost ruined the harvest of the fruit farms that ____ .
12. John met the supervisor of the employees who ____ .
13. The social worker greeted the nurse of the senior-citizens who ____ .
14. We were amused at the articles of the newspaper that ____ .
15. The insurance company covered the furniture of the apartments that ____ .
16. The hacker attacked the web sites of the service provider that ____ .
17. A stranger blackmailed the butler of the royals who ____
18. The scientist criticised the method of the studies that ____ .
19. The secret service confiscated all files of the organisation that ____ .
20. The assassin saw the bodyguard of the diplomats who ____ .
21. The astronomer observed the stars of the spiral galaxy that ____ .
22. The homeowner kept the letters of the estate agency that ____ .
23. The porter smiled at the children of the hotel resident who ____ .
24. The scholar studied the language of the tribes that ____ .