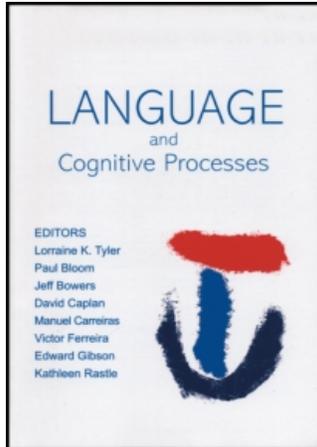


This article was downloaded by:[Max Planck Inst & Research Groups Consortium]  
On: 9 November 2007  
Access Details: [subscription number 769611911]  
Publisher: Psychology Press  
Informa Ltd Registered in England and Wales Registered Number: 1072954  
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Language and Cognitive Processes

Publication details, including instructions for authors and subscription information:  
<http://www.informaworld.com/smpp/title~content=t713683153>

### Relative clause attachment in Dutch: On-line comprehension corresponds to corpus frequencies when lexical variables are taken into account

Timothy Desmet<sup>a</sup>, Constantijn De Baecke<sup>a</sup>, Denis Drieghe<sup>a</sup>, Marc Brysbaert<sup>b</sup>,  
Wietske Vonk<sup>c</sup>

<sup>a</sup> Ghent University, Ghent, Belgium

<sup>b</sup> Royal Holloway, University of London, UK

<sup>c</sup> Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Online Publication Date: 01 June 2006

To cite this Article: Desmet, Timothy, De Baecke, Constantijn, Drieghe, Denis,  
Brysbaert, Marc and Vonk, Wietske (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

To link to this article: DOI: 10.1080/01690960400023485

URL: <http://dx.doi.org/10.1080/01690960400023485>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## **Relative clause attachment in Dutch: On-line comprehension corresponds to corpus frequencies when lexical variables are taken into account**

Timothy Desmet, Constantijn De Baecke, and  
Denis Drieghe  
*Ghent University, Ghent, Belgium*

Marc Brysbaert  
*Royal Holloway, University of London, UK*

Wietske Vonk  
*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

Desmet, Brysbaert, and De Baecke (2002a) showed that the production of relative clauses following two potential attachment hosts (e.g., ‘Someone shot the servant of the actress who was on the balcony’) was influenced by the animacy of the first host. These results were important because they refuted evidence from Dutch against experience-based accounts of syntactic ambiguity resolution, such as the tuning hypothesis. However, Desmet et al. did not provide direct evidence in favour of tuning, because their study focused on production and did not include reading experiments. In the present paper this line of research was extended. A corpus analysis and an eye-tracking experiment revealed that when taking into account lexical properties of the NP host sites (i.e., animacy and concreteness) the frequency pattern and the on-line comprehension of the relative clause attachment ambiguity do correspond. The implications for exposure-based accounts of sentence processing are discussed.

---

Correspondence should be addressed to Timothy Desmet, Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium. Email: Timothy.Desmet@ugent.be

We acknowledge the support of the Research Foundation – Flanders (FWO – Vlaanderen) of which Timothy Desmet is a Postdoctoral Fellow and Denis Drieghe a Research Assistant. This research was also supported in part by grant G.0250.99 of the Research Programme of the Research Foundation – Flanders to Constantijn De Baecke.

## INTRODUCTION

In the last decade, the possibility of experience-based sentence parsing (as opposed to principle-based parsing) has gained ground in psycholinguistics. Indicative of this trend is the appearance of influential probabilistic models of sentence parsing (e.g., Crocker & Brants, 2000; Jurafsky, 1996; Sturt, Costa, Lombardo, & Frasconi, 2003) and neural networks that are capable of learning grammatical patterns on the basis of previous exposure (e.g., Altmann, 2002; Rohde, 2002; Tabor, Juliano, & Tanenhaus, 1997). In addition, numerous behavioural studies have been published that started from corpus data to investigate whether the most frequent structure is also the easiest to process in sentences with local syntactic ambiguities (e.g., Desmet, Brysbaert, & De Baecke, 2002a; Desmet & Gibson, 2003; Gibson & Schütze, 1999; Igoa, Carreiras, & Meseguer, 1998; Mak, Vonk, & Schriefers, 2002; Mitchell & Brysbaert, 1998).

One of the earliest exposure-based models of syntactic ambiguity resolution was the tuning hypothesis, proposed by Mitchell and colleagues (Brysbaert & Mitchell, 1996; Mitchell & Cuetos, 1991; Mitchell, Cuetos, Corley, & Brysbaert, 1995). This view claims that the human sentence parser is experience-based and that the initial parsing choices in syntactic ambiguity resolution are made on the basis of the relative frequencies with which the reader or listener has resolved the syntactic ambiguity in the past. According to the hypothesis, there will always be an initial bias towards the structural interpretation that occurs most frequently in the language. This model was proposed to explain cross-linguistic differences in the attachment of relative clauses in sentences like (1).

(1) Someone shot the servant of the actress who was on the balcony.

The syntactic ambiguity in this type of sentences (introduced by Cuetos & Mitchell, 1988) involves the fact that the relative clause (RC) can be attached to two possible noun phrases. In the first interpretation, the RC 'who was on the balcony' is attached to the first noun phrase (NP1) 'the servant', meaning that the servant was standing on the balcony. This attachment is commonly referred to as high attachment. The other possible interpretation says that the RC is attached to the second noun phrase (NP2) 'the actress' and is called low attachment.

Probably the most interesting finding about the syntactic ambiguity in (1) is that the preferred interpretation differs across languages, with English preferring low attachment, and many other languages (Dutch, French, German, Spanish) preferring high attachment (for an overview, see Mitchell & Brysbaert, 1998). In line with the tuning hypothesis, evidence has been obtained that in English text corpora low attachment is more prevalent than high attachment, whereas in Spanish and French the

reverse pattern was found (Baltazart & Kister, 1995; Corley, 1996; Cuetos et al., 1996; Mitchell & Brysbaert, 1998; Mitchell et al., 1995).

Two studies, however, provided evidence against the tuning hypothesis and other experience-based parsing theories. First, Gibson and Schütze (1999) argued that the resolution of the conjunction of an NP to three possible host sites, as in (2), did not correspond to corpus frequencies that were collected in relation to this ambiguity (Gibson, Schütze, & Salomon, 1996).

- (2) The salesman ignored a customer with a baby with a dirty face and . . .
- (a) a wet diaper (low conjunction)
  - (b) one with a wet diaper (middle conjunction)
  - (c) one with a baby with a wet diaper (high conjunction)

Whereas in the corpus there were more sentences with middle attachments (to ‘a baby’) than with high attachments (to ‘a customer’), in reading tasks participants had less processing problems with high attachments than with middle attachments. In the second study, Mitchell and Brysbaert (1998) analysed a corpus of Dutch newspaper and magazine articles for sentences like (1), and observed that low-attaching relative clauses were twice as frequent as high-attaching relative clauses, despite the finding that in reading studies Dutch-speaking participants consistently preferred high attachment (e.g., Brysbaert & Mitchell, 1996; Desmet, De Baecke, & Brysbaert, 2002b; Mitchell, Brysbaert, Grondelaers, & Swanepoel, 2000).

Contradictions between corpus frequencies and on-line parsing preferences are of crucial importance, because they suggest that syntactic parsing is not experience-based (or at least not completely). For experience-based models, such contradictions have the true status of a rejection of the null-hypothesis (unlike a convergence between corpus frequencies and parsing preferences, which only has the status of a failure to reject the null-hypothesis). Therefore, it is important to understand these contradictions, in order to know how detrimental they are for syntactic parsing models that learn on the basis of the structures they encounter. So, with respect to structure (2), we need to know whether the contradiction between corpus data and reading data reported by Gibson and Schütze (1999) indeed means that different principles underlie sentence production and sentence reading, as originally thought, or whether some characteristic of the stimulus materials is responsible for the divergent findings. To address this issue, Desmet and Gibson (2003) investigated whether the contradiction could be due to the fact that Gibson and Schütze’s sentences contained the pronoun ‘one’ in the conjunction (i.e., ‘. . . and one with a wet diaper’). Such constructions were very rare in the corpus and showed an attachment

pattern that seemed to deviate from that of the dominant structure, which consisted of full noun phrases throughout, as in (3) and (4).

- (3) A column about a soccer team from the suburbs and an article about a baseball team from the city were published in the Sunday edition. (high conjunction)
- (4) A column about a soccer team from the suburbs and a baseball team from the city was published in the Sunday edition. (middle conjunction)

For these structures without the pronoun ‘one’, Desmet and Gibson (2003) observed that participants had less problems reading sentences like (4) with middle attachment, than sentences like (3) with high attachment, in line with the attachment frequencies in the corpus, suggesting that the contradiction reported by Gibson and Schütze (1999) was less of a problem for experience-based parsing models than thought at first.<sup>1</sup>

Desmet et al. (2002a) wondered whether a similar characteristic in the Dutch stimulus materials could be responsible for the contradiction between the corpus data and the reading data in structure (1). They reanalysed the corpus data presented by Mitchell and Brysbaert (1998), and discovered that when the corpus counts were analysed as a function of a specific lexical property of the attachment sites, there was a level of analysis at which the corpus frequencies agreed with the NP1 bias in the comprehension data. Instead of only looking at the total numbers of RCs that were attached to either NP, Desmet et al. additionally coded the animacy of the NPs.<sup>2</sup> This led to four head types: (1) an animate NP1 and NP2 (e.g., ‘the servant of the actress’), (2) an animate NP1 and an inanimate NP2 (e.g., ‘the author of the novel’), (3) an inanimate NP1 and an animate NP2 (e.g., ‘the car of the salesman’), and (4) an inanimate NP1 and NP2 (e.g., ‘the abstract of the article’). At this level of analysis, it was shown that the overall higher frequency of NP2 attachments in the corpus was exclusively due to the sentences with an inanimate NP1 (types 3 and 4). For the other two types (1 and 2), NP1 attachments were more frequent

---

<sup>1</sup> Further corpus analyses indicated that the high preference as in (2) was also present in the corpus when the sentences were limited to those with a pronoun in the NP-PP-PP construction.

<sup>2</sup> Desmet et al. (2002a) worked with the distinction ‘human / non-human’ rather than with the broader distinction ‘animate / inanimate’, which is used more generally, and which we will adopt here as well. This change of terminology has no implications for Desmet et al. (2002a), as none of their sentences referred to animals, so that all human NPs were also animate and all non-human NPs were inanimate. Furthermore, in an unpublished eye-tracking experiment in our lab we found no difference between attachment preferences between NPs that referred to human entities and NPs that referred to animals.

than NP2 attachments. Interestingly, the items used in the Dutch reading studies (Brysbaert & Mitchell, 1996; Desmet et al., 2002b; Mitchell et al., 2000; Wijnen, 1998) mainly consisted of the last two types of sentences (1 and 2). This means that the animacy of NP1 could be responsible for the divergence between sentence writing and sentence reading. As a first test of this hypothesis, Desmet et al. (2002a) asked participants to write continuations for sentences that differed in the animacy of NP1 and NP2. Participants were given the beginning of a sentence (e.g., ‘Someone shot the servant of the actress who...’) and had to write down the first continuation that came to mind. In line with the corpus data, Desmet et al. found an NP1 attachment preference for sentences with an animate NP1 (‘the servant of the actress’, and ‘the author of the novel’) and an NP2 attachment preference for sentences with an inanimate NP1 (‘the car of the salesman’ and ‘the abstract of the article’). Animacy of NP2 had no statistically significant effect on the attachment preference, although there was a small trend towards fewer attachments to inanimate NP2s as well.

The findings of Desmet et al. (2002a) strongly suggest that the contradiction between the corpus frequencies and the reading preferences observed by Mitchell and Brysbaert (1998) need not be evidence against experience-based theories of syntactic parsing (although they are evidence against the coarse-grain version of the tuning hypothesis; see the General Discussion). However, the study was limited to RC attachment in sentence production (corpus data and sentence completion). This leaves open the question whether a similar pattern will be found in sentence reading.

In the present study, we directly address the contradiction reported by Mitchell and Brysbaert (1998) by comparing corpus data with sentence reading preferences. In addition, we aimed to further our understanding of why animacy is such an important variable in RC attachment.

First, we present data from a new corpus analysis that extended the previous findings to less formal language registers, and that led us to discover another variable that affects RC attachments. Second, we ran an eye-tracking experiment that studied the influences of animacy and the new variable in sentence comprehension.

## CORPUS ANALYSIS

The first goal of this corpus analysis was to make sure that the contradiction between the frequencies of RC attachments in Dutch texts and the participants’ preferences in sentence reading (Desmet et al., 2002a; Mitchell & Brysbaert, 1998) was not due to the fact that the texts were based on articles in newspapers and magazines. In general, these articles use a rather formal language and are corrected by text editors. As these features may lead to stylistic deviations from the more frequent, informal

(spoken) language, we considered it necessary to discard language register as a possible origin of the contradiction. In addition, the new corpus analysis allowed us to investigate the influence of the animacy of the noun phrases in less formal language. Finally, as will be outlined in the Results section, the new corpus analysis also made us sensitive to a new variable that affects RC attachment frequencies to complex noun phrases.

## Method

*Materials.* The counts that we present in this corpus analysis are based on six text registers that can be divided into three types: (1) edited written text, i.e., written texts that were published in newspapers or magazines and that were corrected by a professional editor, (2) unedited written text, i.e., written texts that were not revised by an editor, and (3) written spoken text, i.e., texts obtained from an on-line chat channel. For each of these three types of text we collected a sample from the northern half of Belgium<sup>3</sup> and a sample from the Netherlands, leading to six text registers.

The Belgian sample of edited written text consisted of articles from 'Knack' and 'Het Nieuwsblad'. From 'Knack', a general weekly news-magazine, we included the articles from the first five issues of 1993 (January 7, 14, 21, and 28, and February 4) and 1996 (January 3, 10, 17, 24, and 31), with a total of 700 articles. 'Het Nieuwsblad' is a newspaper, which maintains a website with an electronic text archive. This archive is updated every day with a selection of 4 articles that appeared in the most recent newspaper edition. We included all 546 articles starting from July 14, 1999 until December 30, 1999. The Dutch sample of edited written text came from the text archive of the Dutch newspaper 'De Volkskrant'. All articles from the first 10 internet editions in February 2001 were included, with a total of 446 texts.

For the sample of unedited written text we made use of the Usenet files from the CONDIV-corpus (Grondelaers, Deygers, Van Aken, Van Den Heede, & Speelman, 2000). The internet module Usenet is used to debate off-line and asynchronously in a number of newsgroups. The members of the newsgroup express their opinion on a given topic by sending emails that are appended to a thread of previous messages on the same topic. All Belgian and Dutch Usenet files from the CONDIV-corpus (consisting of respectively 4,980,780 and 7,748,436 words) were added to our corpus.

---

<sup>3</sup> There is some confusion about whether the language used in the northern part of Belgium should be called Flemish or Dutch, because there are quite large differences between the spoken regional dialects and the standard language taught in school and used for official communication. We will use the term Dutch, because there are no written representations of the regional Flemish dialects and because the standard written language is the same in the Netherlands and in the northern part of Belgium.

Finally, we included all Internet Relay Chat (IRC) files from the same CONDIV-corpus. IRC is an internet module that permits people to communicate on-line and synchronously via chat channels. Because IRC-users try to adapt their written communication to the principles of spoken conversations, the materials that stem from this source show numerous characteristics of spoken language. Therefore, Grondelaers et al. (2000) define it as 'written spoken Dutch'. All IRC files from the CONDIV-corpus were enclosed in our corpus. The Belgian sample contained 8,207,007 words and the Dutch equivalent 6,965,291 words.

*Procedure.* In order to find the critical sentences in our text files, we used a concordance program (Concapp Version 3.0 for Windows 95) that allowed us to extract all sentences that contained the word 'die'. Next, we examined the extracted pool of sentences and sorted out all instances in which 'die' was a relative pronoun that referred to a complex head with the NP1-van-NP2 structure. Subsequently, it was decided whether an NP1 or an NP2 attachment was made. Instances that could not be disambiguated by means of a semantic or syntactic cue were excluded from the corpus. In order to have local ambiguities with real discourse entities as candidates, only those instances with a referential NP1 and a referential NP2 were included. As a consequence singular NPs that were not introduced by a determiner were excluded. For the remaining instances we categorised the animacy (animate or inanimate) of NP1 and NP2. This was done by three independent judges. When the judgements diverged, the categorisation was decided by deliberation.

## Results

The numbers of high and low attachments we obtained for each of the six text registers are presented in Table 1. In line with previous corpus studies in Dutch (Desmet et al., 2002a; Mitchell & Brysbaert, 1998), the majority of local RC attachment ambiguities were disambiguated in favour of an NP2 interpretation (773 out of 1065 instances, i.e., 73%). Table 1 shows that this pattern is present for all six text registers that we used.

As was the case for the study of Desmet et al. (2002a), the results in Table 2 show that the overall NP2 bias in the corpus is entirely due to those instances that contain an inanimate NP1 (707 out of 863 instances, i.e., 82%). The instances with an animate NP1 were more frequently disambiguated towards the NP1 interpretation (136 out of 202 instances, i.e., 67%). The NP1 bias was also slightly larger for inanimate NP2s than for animate NP2s (78% vs. 63% for an animate NP1, and 21% vs. 16% for an inanimate NP1).

TABLE 1  
Number and percentages of NP1 and NP2 attachments in the corpus for each of the six text registers

<i>Type</i>	<i>Origin</i>	<i>NP1</i>		<i>NP2</i>		<i>Total</i>
		<i>No.</i>	<i>%</i>	<i>No.</i>	<i>%</i>	
Edited	Belgian	92	30	211	70	303
	Dutch	17	18	75	82	92
Unedited	Belgian	77	28	196	72	273
	Dutch	67	26	187	74	254
Spoken	Belgian	25	38	41	62	66
	Dutch	14	18	63	82	77
Total		292	27	773	73	1,065

While we were scoring the different nouns as animate or inanimate, we noticed that not all nouns referred to the concrete, highly imaginable entities we spontaneously associate with these categories (i.e., individuals for animate nouns; and tangible objects for inanimate nouns). Quite often, the nouns referred to rather abstract notions such as 'government' and 'trade union' for the animate category, and 'performance' and 'vision' for the inanimate category. In addition, we got the impression that the attachment frequencies were influenced by the concreteness of the noun, in particular when the noun referred to an animate entity. To examine this impression, we coded all stimuli for concreteness as well, the result of which is shown in Table 3.

It is clear from Table 3 that the higher frequency of NP1 attachments with animate NP1s was entirely due to those sentences in which NP1

TABLE 2<sup>a</sup>  
Number of NP1 and NP2 attachments for each of the four head types obtained by crossing animacy (animate vs. inanimate) and attachment site (NP1 vs. NP2)

<i>Type of NP1</i>	<i>Type of NP2</i>		<i>Total</i>
	<i>Animate</i>	<i>Inanimate</i>	
Animate	94-54	42-12	136-66
Inanimate	82-435	74-272	156-707
Total	176-489	116-284	292-773

<sup>a</sup> In each cell of the table, the first-mentioned number is the number of NP1 attachments and the second-mentioned number is the number of NP2 attachments.

TABLE 3<sup>a</sup>

Number of NP1 and NP2 attachments for each of the 16 head types obtained by crossing animacy (animate vs. inanimate), concreteness (concrete vs. abstract), and attachment site (NP1 vs. NP2)

<i>Type of NP1</i>		<i>Type of NP2</i>				<i>Total</i>
		<i>Animate</i>		<i>Inanimate</i>		
		<i>Concrete</i>	<i>Abstract</i>	<i>Concrete</i>	<i>Abstract</i>	
Animate	Concrete	19–10	56–14	14–4	22–8	111–36
	Abstract	<b>8–28</b>	11–2	1–0	5–0	25–30
Inanimate	Concrete	16–69	<b>19–11</b>	17–26	7–18	59–124
	Abstract	25–290	22–65	16–91	34–137	97–583
Total		68–397	108–92	48–121	68–163	292–773

<sup>a</sup> In each cell of the table, the first-mentioned number is the number of NP1 attachments and the second-mentioned number is the number of NP2 attachments.

referred to a concrete being (concrete: 76% NP1 attachment; abstract: 45% NP1 attachment), and that the higher frequency of NP2 attachments with inanimate NP1s is especially pronounced when NP1 is abstract (concrete: 68% NP2 and abstract: 86% NP2). As a matter of fact, two combinations of concrete and abstract nouns yielded a pattern that was opposite to the overall pattern revealed in Table 2 (see the bold cells in Table 3). There was an NP2 advantage when NP1 referred to an abstract animate entity and NP2 to a concrete animate entity (e.g., ‘the football club of the trainer’); and there was an NP1 advantage when NP1 referred to a concrete inanimate entity and NP2 to an abstract animate entity (e.g., ‘the report of the committee’).

## Discussion

The results of this corpus study replicated the two major findings for Dutch relative clause attachment reported by Desmet et al. (2002a). First, the overall NP2 attachment preference was replicated in all of the six text registers that we sampled. Second, it was shown that when the animacy of NP1 was taken into account, there was clear interaction between the animacy of NP1 and the attachment of the relative clause. When NP1 was animate, RCs were predominantly attached to this noun phrase; when it was inanimate, the majority of RCs modified NP2.

In addition, the new corpus study extended our knowledge in two ways. First, we ascertained that the distribution of RC attachments generalises to different language registers of Dutch. The overall NP2 attachment bias and

the interaction with animacy were not only present in edited newspaper and magazine articles, but also in unedited text and text generated through chat channels. This suggests that the pattern of results is a fundamental characteristic of the whole language.

Second, we found that the influence of animacy is particularly true for nouns that refer to concrete animate beings (i.e., specific humans or animals). Nouns that refer to abstract animate entities ('government', 'trade union', 'board', 'club', 'organization') were less likely to be modified by a RC. This was particularly true when the other noun of the complex head did refer to a concrete entity. So, the high attachment bias for animate NP1s was not found when NP1 referred to an abstract animate entity and NP2 to a concrete animate entity (e.g., 'the reading group of the teacher'; see Table 3). Similarly, the low attachment bias for inanimate NP1s was not found when NP1 referred to a concrete object and NP2 to an abstract animate entity (e.g., 'the books of the reading group'). Implications of these findings for our understanding of RC attachment will be discussed in the General Discussion. First, we examined whether a similar pattern is found in on-line sentence reading.

### Eye-tracking experiment

Thus far, nearly all experimental evidence related to structure (1) has been based on stimulus materials that contained heads of the type concrete animate (human) NP1 and concrete animate (human) NP2 (e.g., 'the servant of the actress', 'the daughter of the colonel'). There are two main reasons for this selection. First, in English it is difficult to combine animate and inanimate noun phrases because one never knows how strongly participants expect the relative pronoun 'who' to be used for animate entities (e.g., must it be 'the author of the book who came to town' rather than 'the author of the book that came to town?'). Second, the use of animate beings allowed researchers to easily solve the local ambiguity created by the RC attachment, for instance by capitalising on the gender of the persons introduced by NP1 and NP2 (e.g., 'the servant of the actress who had *his/her* arm in a cast'). However, as shown in Table 3, these constructions form but a tiny segment of all sentences with this particular structure that are produced in a language (i.e., 29/1065, or less than 3%), and at least in Dutch induce a different RC attachment bias (66% NP1) than the overall attachment bias (27% NP1).

A much richer picture of the correspondences between sentence reading and sentence writing can be obtained by looking at the complete first column of Table 3. What this column suggests, is that we should find fewer reading difficulties when the RC is attached low for three out of the four combinations of NP1 and NP2. Only for one combination would we find

the reverse pattern, namely when a concrete animate being is paired to another concrete animate being. Or to phrase it more poignantly: If there is a perfect correlation between sentence perception and sentence production, for the first column of Table 3, the high attachment preference, that has been so robust in previous sentence reading research, would be limited to only one out of four conditions. This is the task we set ourselves in the present experiment.

## Method

*Participants.* A total of 48 undergraduate students of Ghent University participated individually for course credit. All participants had normal vision or wore contact lenses. They were all native speakers of Dutch and unaware of the goal of the study.

*Materials and design.* Thirty-two sets of eight sentences were constructed. The eight sentences in a set were obtained by crossing the type of NP1 (animate concrete, animate abstract, inanimate concrete, and inanimate abstract) and the attachment of the RC (high [NP1] vs. low [NP2]). All NP2s referred to concrete animate beings (which stayed the same within a set). An example set of sentences is given in (4).

(4a) inanimate, abstract NP1

De bevolking zonder toekomstperspectieven respecteert **de beslissingen** van de president die (*garanderen / garandeert*) dat er geen oorlog komt. [The population without any future perspectives respects **the decisions** of the president that (*guarantee / guarantees*) there will be no war.]

(4b) inanimate, concrete NP1

De bevolking zonder toekomstperspectieven respecteert **de documenten** van de president die (*garanderen / garandeert*) dat er geen oorlog komt. [The population without any future perspectives respects **the documents** of the president that (*guarantee / guarantees*) there will be no war.]

(4c) animate, abstract NP1

De bevolking zonder toekomstperspectieven respecteert **de organisaties** van de president die (*garanderen / garandeert*) dat er geen oorlog komt. [The population without any future perspectives respects **the organizations** of the president that (*guarantee / guarantees*) there will be no war.]

(4d) animate, concrete NP1

De bevolking zonder toekomstperspectieven respecteert **de raadgevers** van de president die (*garanderen / garandeert*) dat er geen oorlog komt. [The population without any future perspectives respects **the advisors** of the president that (*guarantee / guarantees*) there will be no war.]

The subject NP and the main verb of the sentence always preceded the complex NP1-van-NP2 head, which was the object of the sentence. All NP1s and NP2s were words that took 'de' as determiner, so that the relative pronoun 'die' could refer to both NP1 and NP2. The syntactic ambiguity was resolved by the number of the verb within the relative clause. Half of the items contained a plural NP1 and a singular NP2, while the other half had a singular NP1 and a plural NP2. Using this type of disambiguation allowed us to match the disambiguation for length, because we used Dutch verbs that had the same length in their plural and singular form ('garanderen' [guarantee] vs. 'garandeert' [guarantees]). The NP1s in the different conditions did not differ in length or frequency (all  $t < 1.68$ , all  $p > .10$ ). All items can be found in Appendix A. In addition to the 32 experimental sentences, 88 filler sentences were used. Because of a programming error, not all sentences of set 16 were presented to the participants. Therefore, this set was excluded from all the analyses reported below.

*Procedure.* Participant's eye movements were recorded by an SMI Eyelink headband-mounted eye-tracking system. The Eyelink system samples both the horizontal and vertical signal every 4 ms and is based on an infrared video-based tracking technology that happens simultaneously for both eyes. Although the Eyelink system compensates for head position, this compensation is not accurate enough to allow single character resolution. Therefore, we installed a height-adjustable chin rest at a fixed distance (75 cm) from the stimulus display.

Participants were asked to put their head on the chin rest and to move as little as possible. A practice session preceded the experimental session to allow participants to become familiar with the eye-tracking equipment and the experimental procedure. Both the practice session and the experimental session started with a calibration and validation procedure. In the calibration procedure the participants were asked to fixate nine calibration points that were presented randomly one at the time in the form of a 9-point grid. The calibration was evaluated by a built-in routine and each eye's calibration was graded 'good', 'poor', or 'failed'. Only when the calibration of both eyes was graded 'good' was the validation procedure started. The validation procedure assessed the accuracy of the system in predicting gaze position from pupil position. In the validation phase, the same nine target points were presented as in the calibration procedure. When the participants fixated these, the calibration values were used to estimate the gaze position of the participant and to calculate the error (i.e., the difference between the target position and the computed gaze position). As in the calibration procedure, each eye was graded separately and was accepted only when the maximal distance between the target

position and the computed gaze position did not exceed 0.5 for each of the nine target points.

After the calibration and validation procedures were completed, the sentences were presented in a different random order for each participant. Each trial started with a calibration check (a single fixation point in the centre of the screen) and was adjusted in case the check was negative. Participants were asked to read each sentence as soon as it was presented and to push a button when they had finished. The experimental items were presented on two or three lines. The first line contained the sentence beginning up to the main verb of the sentence. The second line started with the NP1 so that the critical region (the disambiguating verb) was always presented in the middle of the second line. Only in those sentences with a long RC was a third presentation line needed. In order to encourage participants to read for meaning, they were informed that occasionally a simple yes–no question would be asked about the sentence they had read (30 of the 120 sentences). None of the questions was about the research question (i.e., about the attachment of the RC). The experimenter told them whether they had answered the question correctly or not. The experiment started with a practice session consisting of eight practice sentences, two of which were followed by a question. The entire experiment took about 40 minutes.

## Results

For analysis purposes the target sentences were divided into seven regions, illustrated in (5). Region 1 was the beginning of the sentence up to NP1. Region 2 consisted of the NP1. Region 3 was the prepositional phrase containing the preposition ‘van [of]’ and NP2. Region 4 contained the relative pronoun ‘die’. Region 5 contained the disambiguating verb. Region 6 contained the following two words. Finally, Region 7 consisted of the remainder of the sentence. We ran ANOVAs with two repeated measures (NP1 type and attachment site) on each of the seven regions. These analyses were done both over participants ( $F_1$ ) and over items ( $F_2$ ). Here, we will concentrate on results for the regions from the disambiguating verb on (Regions 5, 6, and 7). Comparing reading times for the previous regions is not very insightful because they contain different words in the different conditions. Moreover, they cannot reflect anything concerning attachment preferences. To illustrate that the effects on the disambiguating region do not simply reflect spill-over from the prior regions the means of these regions will be presented in the tables and the results of the analyses on these previous regions (Region 1 to 4) can be found in Appendix B.

(5) De bevolking zonder toekomstperspectieven respecteert / de beslissingen / van de president / die / garanderen / dat er / geen oorlog komt. [The population without any future perspectives respects / the decisions / of the president / that / guarantee / that there / will be no war].

*Cumulative Region Reading Time (CRRT).* We started our analyses by calculating the mean CRRT for each of the seven regions (see Table 4). CRRT is defined as the sum of the fixations between the moment when the eyes first cross the front border of the region and the moment when they first cross the back border. The difference between CRRT and first-pass reading time (FPRT) is that regressions originating from a particular region are added to the CRRT of that region, but they are not added to the FPRT. It has been argued that CRRTs are very sensitive to parsing difficulties (e.g., Brysbaert & Mitchell, 1996; Liversedge, Paterson, & Pickering, 1998) because processing difficulties manifest themselves either by prolonged reading of the disambiguating region or by rereading the previous ambiguous part of the sentence.

The analyses on the disambiguating region (Region 5) showed a significant main effect of NP1 type,  $F_1(3,138) = 4.71, p < .01$ ;  $F_2(3,90) = 9.09, p < .001$ . The animate concrete condition was read more slowly than the other conditions. There was no main effect of attachment site ( $F_1$  and  $F_2 < 1$ ). Most importantly, as predicted, the analysis on Region 5 revealed

TABLE 4  
Mean cumulative region reading times (CRRTs, in milliseconds) in the eye-tracking experiment for each of the seven regions as a function of head type and attachment site

	<i>Regions</i>						
	<i>1</i> <i>begin</i>	<i>2</i> <i>NP1</i>	<i>3</i> <i>PP</i>	<i>4</i> <i>die</i>	<i>5</i> <i>V</i>	<i>6</i> <i>next 2</i>	<i>7</i> <i>rest</i>
Animate-Concrete							
NP1	1788	386	757	83	515	216	2383
NP2	1811	365	690	77	656	253	2506
Animate-Abstract							
NP1	1918	418	629	53	459	228	2210
NP2	1760	419	574	57	434	195	2191
Inanimate-Concrete							
NP1	1797	361	630	81	487	249	2210
NP2	1822	380	618	50	464	228	2091
Inanimate-Abstract							
NP1	1748	387	524	50	489	246	2283
NP2	1816	459	553	38	447	218	2232

a significant interaction between NP1 type and attachment site,  $F_1(3,138) = 3.40, p < .05$ ;  $F_2(3,90) = 3.76, p < .05$ . The numerical pattern was completely identical to that of the corpus frequencies (see Table 3, first column). When NP1 was animate concrete, NP2 attachments needed more time to be processed than NP1 attachments (656 vs. 515 ms). In contrast, for the three other combinations, NP1 attachments took more time to process than NP2 attachments (459 vs. 434 ms for animate-abstract NP1s, 487 vs. 464 ms for inanimate-concrete NP1s, and 489 vs. 447 ms for inanimate-abstract NP1s). However, planned comparisons revealed that only the NP1 bias in the animate concrete condition was significant by itself,  $F_1(1,46) = 6.31, p < .05$ ;  $F_2(1,30) = 5.51, p < .05$ .

Part of the effect due to attachment site spilled over to Region 6, the first region following the disambiguating verb. That is, there was a perfect correlation of the attachment site differences described for Region 5 and those observed in Region 6. However, none of the effects was significant when the ANOVA was confined to Region 6 (all  $F < 1$ ).

At the end of the sentence (Region 7) the main effect of NP1 type reappeared in the analysis over participants,  $F_1(3,138) = 3.31, p < .05$ , but not in the analysis over items,  $F_2(3,90) = 1.52, p = .22$ . The animate concrete condition was read more slowly than the other three conditions,  $F_1(1,46) = 5.76, p < .05$ ;  $F_2(1,30) = 4.19, p < .05$ , which did not differ from each other (all  $F < 1$ ). The main effect of attachment site and the interaction were not significant (all  $F < 1$ ).

*First-Pass Reading Time (FPRT).* The predicted interaction in the CRRTs in the disambiguating region could be due to differences in first-pass reading or to differences in the number of regressive eye-movements. Therefore we also calculated FPRTs and percentage of regressions. FPRT was defined as the sum of fixations between the moment the eyes first entered the region and the moment they first left the region either to the left or the right. Mean FPRTs for each of the seven regions are presented in Table 5.

The most interesting question is whether the significant interaction on CRRTs at Region 5 is due to first-pass reading. This was not the case: the ANOVA on the FPRTs showed that there were no significant effects at all on the disambiguating region (all  $F < 1$ ). Also the ANOVAs on the two final regions (Region 6 and 7) revealed no significant effects in first-pass reading (all  $F < 1.59$ , all  $p > .21$ ).

*Percentage of regressions.* Given that the interaction in CRRTs was not due to first-pass reading times we further calculated the percentage of first-pass regressions, to see whether the interaction was present here (see Table 6). We defined percentage of regressions as the number of trials in

TABLE 5

Mean first-pass reading times (FPRTs, in milliseconds) in the eye-tracking experiment for each of the seven regions as a function of head type and attachment site

	<i>Regions</i>						
	<i>1</i> <i>begin</i>	<i>2</i> <i>NP1</i>	<i>3</i> <i>PP</i>	<i>4</i> <i>die</i>	<i>5</i> <i>V</i>	<i>6</i> <i>next 2</i>	<i>7</i> <i>rest</i>
Animate-Concrete							
NP1	1788	351	482	55	370	165	884
NP2	1811	344	479	58	381	167	772
Animate-Abstract							
NP1	1918	391	437	43	388	180	862
NP2	1760	387	444	39	367	156	827
Inanimate-Concrete							
NP1	1797	322	413	53	370	187	809
NP2	1822	355	440	31	386	165	811
Inanimate-Abstract							
NP1	1748	381	425	45	376	171	807
NP2	1816	379	443	31	373	177	833

TABLE 6

Mean percentage of first-pass regressions in the eye-tracking experiment for each of the seven regions as a function of head type and attachment site

	<i>Regions</i>						
	<i>1</i> <i>begin</i>	<i>2</i> <i>NP1</i>	<i>3</i> <i>PP</i>	<i>4</i> <i>die</i>	<i>5</i> <i>V</i>	<i>6</i> <i>next 2</i>	<i>7</i> <i>rest</i>
Animate-Concrete							
NP1	0	4	28	–	16	12	70
NP2	0	4	25	–	25	13	76
Animate-Abstract							
NP1	0	2	21	–	11	14	70
NP2	0	4	18	–	14	13	65
Inanimate-Concrete							
NP1	0	5	26	–	16	13	70
NP2	0	2	22	–	11	15	71
Inanimate-Abstract							
NP1	0	1	12	–	13	7	71
NP2	0	5	15	–	11	11	72

which the eyes leave a region to the left, relative to the number of trials this region has been looked at during first-pass reading.

The analysis on Region 5 revealed that the significant interaction observed in CRRTs, was due to the percentage of regressions participants made from this region,  $F_1(3,138) = 2.92, p < .05$ ;  $F_2(3,90) = 3.23, p < .05$ . Also the main effect of NP1 type that was present in CRRTs showed up in the analysis on percentage of regressions, even though it was only fully significant in the analysis over items,  $F_1(3,138) = 2.35, p = .08$ ;  $F_2(3,90) = 3.61, p < .05$ . Again, at the last two regions there were no significant effects in the percentage of regressions (all  $F < 1.68, p > .17$ ).

*Correlations between reading time measures and corpus frequencies.* Even though the CRRT and percentage of regressions revealed a significant interaction between head type and attachment preference on the disambiguating region and even though the numerical pattern mirrored the corpus frequencies exactly, the planned comparisons (NP1 vs. NP2 attachment) were only significant for the animate concrete NP1s. In order to further investigate whether the reading time data were statistically in line with the corpus data we decided to look at a number of correlations. First, for each of the 48 participants we calculated a correlation between the NP1 reading time advantage in the four head type conditions (the mean NP2 reading time minus the mean NP1 reading time for each head type) and the corresponding corpus bias. We created two versions of this corpus bias: a general corpus bias was calculated over all types of NP2 and was based on the last column of Table 3: animate concrete NP1s showed a 76% NP1 bias (111/147), animate abstract NP1s a 45% NP1 bias (25/55), inanimate concrete NP1s a 32% NP1 bias (59/183), and inanimate abstract NP1s a 14% NP1 bias (97/680). A more specific corpus bias was calculated over animate concrete NP2s (the type of NP2 that was used in our experiment) and was based on the first column of Table 3: animate concrete NP1s showed a 66% NP1 bias (19/29), animate abstract NP1s a 22% NP1 bias (8/36), inanimate concrete NP1s a 19% NP1 bias (16/85), and inanimate abstract NP1s a 8% NP1 bias (25/315). Then, these correlations (between NP1 reading time advantage and NP1 corpus bias) were inserted as raw data values into a one-sample t-test to investigate whether the correlations were greater than zero.

These analyses showed that on the disambiguating region there was a significant correlation between reading times and corpus bias for the CRRTs [for the specific corpus bias: mean  $r = .19, t(47) = 2.21, p < .05$ ; for the general corpus bias: mean  $r = .17, t(47) = 2.19, p < .05$ ] and for the percentage of regressions [for the specific corpus bias: mean  $r = .17, t(47) = 2.05, p < .05$ ; for the general corpus bias: mean  $r = .19, t(47) = 2.29, p < .05$ ], but not for the FPRTs [for the specific corpus bias: mean  $r = -.02,$

$t < 1$ ; for the general corpus bias: mean  $r = -.02$ ,  $t < 1$ ]. This pattern of results mirrors the results obtained by the ANOVAs above and confirms that the significant interactions that were found are indicative of an alignment between reading times and corpus frequencies.

*Plausibility check.* Another worry was that the reading time differences might not reflect attachment preferences, but were simply due to the fact that some head types were more plausible subjects of the RC than other head types. It has been found that animate noun phrases are more plausible subjects than inanimate noun phrases (e.g., Bock, Loebell, & Morey, 1992). Looking at (4) it could be that ‘the decisions of the president’ are a less plausible subject for ‘guarantee’ than ‘the advisors of the president’. Therefore, we rephrased the different RCs and their heads into active sentences and presented them in a plausibility rating study. For instance, (4a) was changed into ‘The decisions of the president guarantee there will be no war’. Similar changes were made for the other three versions of NP1 interpretation (‘the documents of the president’, ‘the organisations of the president’, and ‘the advisors of the president’), and we also included the NP2 interpretation in the rating study (‘The president guarantees there will be no war’).

Twenty-five new participants, who did not take part in the eye-tracking experiment, rated the sentences on a 5-point scale (1 = implausible, 5 = plausible). The five versions of each of the 32 items (4 NP1 versions and 1 NP2 version) were presented in five lists according to a Latin-Square design and were intermixed with 32 filler items (16 plausible and 16 implausible fillers). The results of the plausibility rating study indicated that the animate heads were slightly less plausible than the other heads: animate concrete NP1 (3.12), animate abstract NP1 (3.21), inanimate concrete NP1 (3.43), inanimate abstract NP1 (3.46) and the NP2 version (3.47). This goes against the explanation that differences in plausibility could underlie the interaction we found. So, we can be confident that our findings are not an artifact of a plausibility confound.

*An extra analysis for the sentences with animate-concrete NP1s.* As shown in Tables 4–6, the sentences in which two concrete animate entities (in this case, two humans) were introduced, were the odd ones out. Not only did they lead to a different attachment preference, but they were also more difficult to process, already from the moment the second noun phrase (entity) was introduced (see Appendix B). In total, they took half a second longer to read (6.25 s) than the other three types of sentences (5.75 s). When we constructed these materials, we made sure that not all sentences contained NP1-van-NP2 combinations in which NP2 was a necessary argument of NP1, as in ‘The old baker in the town was envious of the daughters of the millionaire who ...’, where it is next to unacceptable not

to further define the NP1 ‘daughters’ before continuing the sentence (i.e., \* ‘The old baker in the town was envious of the daughters who...’). As a matter of fact, about half of our sentences contained NP1-van-NP2 combinations in which the NP2 could easily be dropped (as in ‘The lawyer with the Sicilian ancestors was fascinated by the customer of the drivers who...’). In this type of sentences, the relationship of the two NPs is better characterized as an adjunct relationship (see Schütze & Gibson, 1999, for further information about the distinction between arguments and adjuncts).

To find out whether there was a distinction between the sentences with an argument relationship between both NPs and those with an adjunct relationship, we divided the sentences with animate concrete NP1s as a function of this distinction (see Appendix A). Table 7 shows the results of the CRRTs for these two types of sentences.

As can be seen in Table 7, there was little difference between both types of sentences. There was an NP1 preference both for adjunct sentences and for the argument sentences (the effect of low vs. high attachment [summed over Regions 5–7] amounted to 315 ms for adjunct sentences, and 288 ms for argument sentences). ANOVAs with the additional independent variable argument/adjunct in the  $F_2$  analysis revealed that in none of the regions there was an interaction between argument/adjunct and attachment preference (Regions 1, 2, 4, 5, 6, and 7:  $F < 1$ ; Region 3:  $F(1,29) = 2.16, p = .15$ ). There was also no main effect of thematic structure on any of the regions (all  $F < 1.27$ , all  $p > .26$ ). The only significant effect we found confirmed the main effect of attachment site in Region 5,  $F(1,29) = 5.40, p < .05$ . In the other regions there was no main effect of attachment site (all  $F_s < 1$ ). Given these post-hoc analyses, it seems highly unlikely that differences in argument structure could be responsible for the deviant pattern of the sentences with animate concrete NP1s.

TABLE 7

Mean cumulative region reading times (CRRTs, in milliseconds) of the eye-tracking experiment for each of the seven regions as a function of relationship between NP1 and NP2 and attachment site (animate concrete NP1s only)

	<i>Regions</i>						
	<i>1</i> <i>begin</i>	<i>2</i> <i>NP1</i>	<i>3</i> <i>PP</i>	<i>4</i> <i>die</i>	<i>5</i> <i>V</i>	<i>6</i> <i>next 2</i>	<i>7</i> <i>rest</i>
Argument relationship							
NP1	1807	388	813	78	569	233	2432
NP2	1788	375	645	84	698	285	2539
Adjunct relationship							
NP1	1776	384	696	88	459	197	2337
NP2	1836	359	735	72	635	214	2459

## Discussion

The main purpose of this experiment was to find out whether the contradiction between the reading data in Dutch (Brybaert & Mitchell, 1996; Desmet et al., 2002b; Mitchell et al., 2000; Wijnen, 1998) and the corpus findings reported by Mitchell and Brybaert (1998) was indeed due to the animacy of the nouns as suggested by Desmet et al. (2002a). In addition, we wanted to know whether there was a distinction between concrete and abstract animate nouns as suggested by Table 3.

The significant interaction between head type and attachment preference and the significant correlations between NP1 reading time advantage and NP1 bias in the corpus convincingly showed the fit between corpus data and reading data. Of the four types of sentences tested, only one induced a preference for high attachment (to NP1), namely the sentences with two animate concrete nouns. This is the type of sentences that has been examined in nearly all previous research, not only in Dutch but also in other languages. For the other three types of sentences, there was a preference for low attachment, as predicted by the corpus data.

One reason for this interaction could be that the influence of concreteness and animacy on the RC attachment preference was an artifact of the thematic relationship between the two noun phrases. For instance, in the construal theory it is argued that the attachment preference in the RC ambiguity is heavily influenced by the argument structure of the complex head containing the two NPs (see p. 73 in Frazier & Clifton, 1996). According to this theory, if NP2 is an argument of NP1, then the RC is associated to the entire 'NP1-of-NP2' structure and both NP1 and NP2 are considered as possible attachment sites. On the other hand, if NP2 is not an argument of NP1, the RC is associated to the 'of-NP2' structure and only NP2 is available as a potential host. Consequently, the NP1 attachment preference will be higher when NP2 is an argument of NP1 than when NP2 is not an argument. Based on this rationale, it could be argued that the significant NP1 bias in the animate concrete condition is due to the fact that two animate concrete entities are predominantly in an argument relation, whereas in the other conditions, the two entities are more often in an adjunct relation. Looking at our items revealed indeed that the conditions other than the animate abstract condition predominantly contained adjunct NP2s (e.g., 'the gym classes of the teacher', 'the dance moves of the ballerina'), although this was not exclusively so (e.g., 'the style of the journalists', 'the intentions of the terrorist'). In contrast, only half of the animate concrete sentences contained an adjunct relation (such as 'the bishop of', 'the soldiers of', 'the doctor of'); the other half had an argument relation (such as 'the daughters of', 'the brother of', 'the boss of'). To test the alternative interpretation that the thematic relationship

between the nouns is the decisive variable, we split the sentences with an animate concrete NP1 into those with an argument structure and those with an adjunct structure (see Table 7). No difference was observed (not even a trend), making it highly unlikely that differences in the thematic relationship between the two possible attachment sites are responsible for the reading data observed.

We also found that the sentences with animate concrete NP1s were more difficult to read than the other types of sentences. This difference started to emerge as soon as NP2 was introduced (Table 4) and was true both for the sentences with an argument structure and those with an adjunct structure (Table 7). The fact that the processing difficulty was also observed in sentences with an argument structure discards the possibility that it is due to the fact that participants did not expect an animate concrete noun to be modified by a PP with another animate concrete noun. In the case of an argument structure, such a modification is precisely expected (e.g., as a continuation of the sentence 'she saw the daughter...'). One possible interpretation could be that the introduction in the discourse representation of two human referents with very similar characteristics causes interference, so that participants have more difficulties keeping the referents apart (Meyer & Bock, 1999). This interpretation would agree with the finding that the extra processing cost largely consisted of regressive eye movements to previous parts of the sentence (Table 6). Whatever the exact interpretation, the findings of our reading study show that the ambiguity in sentence (1) has largely been investigated on the basis of stimulus materials that contain an infrequent and, at least in Dutch, uncharacteristic combination of noun phrases.

It may be important to note that even though the numerical patterns in the eye-tracking study are highly similar to the corpus frequencies, it is strange that the low attachment bias in the inanimate abstract NP1 condition did not reach significance in the planned comparisons, even though the corpus bias is stronger than that in the animate concrete NP1 condition, where the on-line high attachment advantage came out significantly. One probable explanation is that it is impossible to construct sentences for a reading experiment that are in all regards completely representative of the sentences found in the corpus. For instance, in order to keep the variance in reading times as low as possible the head containing the two NPs in our items was always in the object position of the main sentence, the RCs always immediately followed the second NP and they were always subject-extracted RCs. In the corpus, however, the two NPs occupied a range of syntactic positions, often there was linguistic material intervening between the NP2 and the RC, and the structure of the RC was also very diverse. It is at least conceivable that the fact that NP1 was always the object of the main sentence enhanced

the NP1 bias in our reading experiment (increasing the NP1 bias in the animate concrete condition and decreasing the NP2 bias in the inanimate abstract condition). Anyway, even when additional factors slightly influenced the data, it is still clear that the nature of NP1 interacted with the attachment bias in a highly similar way both in comprehension and corpus frequencies.

## GENERAL DISCUSSION

Since Cuetos and Mitchell (1988) presented their initial finding of a cross-linguistic difference between Spanish and English in the attachment of relative clauses to complex heads of the type NP1-of-NP2, sentence (1) has continued to inspire researchers. First, the language difference questioned the then prevailing assumption that the same universal syntactic principles governed parsing in all human languages. Second, the high attachment in Spanish was in disagreement with the widespread view that new incoming information is preferentially attached to the most recent phrase if this does not increase the overall syntactic complexity of the sentence (e.g., the late closure principle in Frazier's (1978) garden-path theory).

Proposals to solve the problems raised by sentence (1) have gone in different directions (see Desmet et al., 2002a for a summary). The direction that concerns us most here, is the one proposed by Mitchell et al. (1995). According to their tuning hypothesis, sentence parsing is not solely based on universal principles (e.g., due to memory limitations), but also depends on the previous experiences of the human parser. Just like people in a tachistoscopic word identification task are more likely to mistakenly report high frequency words for low frequency target words than the other way around, so do syntactic structures with a high frequency in daily use have a priority over syntactic structures with a low frequency in daily use.

Soon afterwards, however, the tuning hypothesis failed on its fourth test on the structure for which it had been formulated. After successful correlations had been obtained between corpus frequencies and reading preferences in Spanish, English, and French, the tuning hypothesis failed for the Dutch language (Mitchell & Brysbaert, 1998). Whereas reading data pointed to a preference for high attachment, there was an overwhelming predominance of low attachments in the corpus (see also Table 1). This contradiction not only questioned the tuning hypothesis, but all other experience-based models of sentence parsing that were proposed around the same time and that have been presented since (see the Introduction for some references), unless a factor could be found that explained the contradiction. The present study (in combination with Desmet et al., 2002a) shows that such a variable exists, and that it has to do with the nature of the nouns in the NP1-of-NP2 head which precedes the

relative clause. In the remainder of this text, we describe what the implications of this finding are for the tuning hypothesis and for experience-based models of sentence parsing in general.

The tuning hypothesis emphasised that structural frequencies need to be taken into account in order to explain human sentence processing behaviour (Mitchell et al., 1995). It strongly argued against purely lexicalist frequency models of sentence processing such as the constraint-based theories that were presented around the same time (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Spivey-Knowlton & Sedivy, 1995; Trueswell, Tanenhaus, & Kello, 1993). Although Mitchell et al. (1995) left open the possibility that mixed accounts (i.e., accounts that take both structural (coarse-grained) and lexical (fine-grained) statistics into account) were compatible with the empirical data available, the first author (Don C. Mitchell) stressed that the coarse-grained version of the tuning hypothesis was consistent with all the evidence presented at that time and that it was a more parsimonious account of parsing data. However, on the basis of the current results it is clear that the coarse-grained version of the tuning hypothesis is no longer tenable and that a mixed version, where both structural and lexical frequencies are counted, is the only viable variant of the tuning hypothesis. What our data show, is that to keep the reading data in line with the corpus biases, features of the words within the structure – such as the animacy and the concreteness of the two nouns – have to be taken into account. For this reason, our data are also problematic for Sturt et al.'s (2003) recent implementation of the tuning hypothesis. In this implementation, the parser at each word tabulates all possible continuations of the syntactic tree, and in cases of multiple possibilities ranks the likelihood of each of them on the basis of the sentence structure processed thus far (this is achieved with a recursive neural network). Sturt et al.'s implementation successfully predicted the low attachment preference for sentences like (1) in English, also when the parser had not encountered this particular structure before (the latter was due to generalisation from other, similar structures). There is no way, however, in which the parser could account for the word-related differences reported here, simply because the current version of the model only takes into account the syntactic categories of the words.

To account for the effects of animacy and concreteness on relative clause attachment in an experience-based model, it is necessary to store this information and to make use of it in on-line parsing decisions. One way to achieve this, would be to encode it at the level of the individual words (i.e., in the word lexicon). Such lexical variables have been incorporated in many models of sentence parsing (e.g., Spivey-Knowlton & Sedivy, 1995; Trueswell et al., 1993; Vosse & Kempen, 2000). However,

a difficulty for this account with respect to sentence (1) is that the RC attachment bias is not in the first place determined by the characteristics of a single word, but by the comparison of two (or more) words. As our eye-tracking experiment showed (Table 4), the same NP2 either attracted a RC attachment or not, as a function of the characteristics of NP1. Similarly, Gibson, Pearlmutter, and Torrens (1999) showed that a large cost was associated with processing the RC when it was attached to the noun 'planet' in the sentence 'The astronomer predicted the orbits of the planet that was observed from the satellite', but not when it was attached to the same noun in the sentence 'The astronomer predicted the changes of the orbits of the planet that was observed from the satellite'. Finally, Mitchell et al. (1995) reported evidence that the same word was the preferred attachment site or not, depending on its position within the head (e.g., 'the doctor of the patients who...' vs. 'the patients of the doctor who...'). So, the attachment decision is determined not only by the characteristics of the words, but also by their position within the sentence.

Given that exposure-based accounts that focus exclusively on either lexical or structural levels are incompatible with the available evidence, the best alternative would be a model that integrates frequency information from different levels of analysis. One example of such an approach is Jurafsky's (1996) probabilistic model in which the disambiguation of potential interpretations is based on conditional probabilities. In this model, the conditional probabilities of the alternative constructions are calculated on the basis of evidence both from syntactic and lexical sources, bottom-up and top-down. To solve the RC attachment ambiguity in (1) and (4), structural as well as lexical information would be taken into account to calculate the probabilities of the two attachment sites. Another example of this approach is Tabor et al.'s (1997) dynamical system, in which simple lexical frequencies and frequencies contingent on an environment of syntactic categories are combined to make predictions of upcoming structures in a recurrent connectionist network.

Another promising approach could be that of McRae, Ferretti, and Amyote (1997). In their view, the thematic roles that phrases play in a sentence ('who does what to whom'), are (partly) based on the features of the words. Some features are more typical for agent roles and others for patient roles. For instance, typical patient features are (Dowty, 1991): <undergoes change of state>, <causally affected by another subject>, <stationary relative to movement of another subject>, and <does not exist independently of the event>. According to McRae et al. (1997), thematic roles are not all or none phenomena, but graded concepts 'formed through the everyday experiences during which people learn about the entities and objects that tend to play certain roles in certain events' (p. 141). One could envisage that the thematic features activated by

NP1 and NP2 (rather than the words themselves) compete with one another to determine which NP is the most likely site to be modified by a RC. This, in combination with structural information, might be able to account for the data of Tables 3 and 4.

The fact that we showed a correlation between sentence production and sentence perception for structure (1) in Dutch puts the experience-based approaches on the map again as a possible explanation of the parsing preferences in this structure. However, an additional challenge is to explain how the corpus counts look the way they do, i.e., why some structures were produced more or less in the first place. For such an explanation we must turn to the sentence production literature. Bock and colleagues have argued that there is a strong tendency to bind animate entities to the subject position of a clause (e.g., Bock, 1986; Bock & Loebell, 1990; Bock, Loebell, & Morey, 1992). When participants are asked to describe pictures that contain animate patients and inanimate agents, they show a bias to form passive sentences, so that they can put the animate entity in the subject position. Similarly, when participants are asked to rate the goodness of sentences, they give higher ratings to sentences with animate subjects than to sentences with inanimate subjects (Corrigan, 1986), and when they are asked to make sentences with a given set of words, they use the animate words of the set more often as the subject of the sentence than the inanimate words (Itagaki & Prideaux, 1985). The general picture seems to be that conceptually more accessible entities (such as animate entities or concrete entities) occupy more important grammatical positions.

The tendency to associate animacy with subjecthood of a sentence may explain why there is such a strong bias to attach a relative clause to an animate noun (Table 3 and 4). Most relative clauses in a language are subject extracted (i.e., the relative pronoun is the subject of the relative clause; see Mak et al., 2002, for corpus evidence on this in Dutch and German). So, there is a strong bias to expect that animate entities in the discourse representation will be the subject (the agent) of an upcoming relative clause. In addition, it has been claimed that animate entities are more accessible in the discourse representation than inanimate entities (e.g., MacDonald, Bock, & Kelly, 1993), which may be a further reason why relative clauses are more likely to be attached to animate nouns.

The impact of concreteness may be understood by taking into account the ideas of McRae et al. (1997), introduced in the previous section. According to these authors, the effects of animacy and conceptual accessibility are not categorical (all-or-none), but are continuous variables dependent on the (thematic) features that are activated by the nouns. There is a huge literature in word recognition and memory research showing that semantic features of concrete words are more rapidly

activated than those of abstract words. Compared to abstract words, concrete words are recognised faster (e.g., Ransdell & Fischler, 1987; Schwanenflugel, Harnishfeger, & Stowe, 1988), recalled better (e.g., Paivio, 1986; Ransdell & Fischler, 1987), and translated faster and more accurately (e.g., de Groot, 1992; de Groot & Hoeks, 1995). So, it does not seem unlikely that the thematic agent role is activated more strongly by a concrete animate noun that refers to an individual living and acting entity (e.g., ‘the colonel’, ‘the actress’), than by an abstract animate noun that refers to an entity which itself is not alive, but which represents a body of individual human beings (e.g., ‘the government’, ‘the board’) that are able to make their own decisions and have several other animate-like features (see Yamamoto, 1999 for further discussion). If this interpretation is right, then we may be able to influence the attachment bias by adding a feature that evokes the thematic agent role, to one of the nouns (McRae et al., 1997). So, we may be able to overcome the NP2 bias for the structure ‘the parish of the priests that’ (‘de parochie van de priesters die’) by using the expression ‘the enterprising parish of the priests that’ (‘de ondernemende parochie van de priesters die’), but not by using the expression ‘the poor parish of the priests that’ (‘de arme parochie van de priesters die’).

## Conclusions

The present studies have established that the only way to understand relative clause attachment in Dutch for sentences like ‘someone shot the servant of the actress who was on the balcony’ is to take into account some characteristics of the words that make up the possible attachment sites. In particular, we have shown that the attachment strongly depends on the animacy of NP1. When NP1 is animate, there is a bias towards high attachment; when NP1 is inanimate there is a bias towards low attachment. (There are only two exceptions to this pattern: Firstly, when both NPs are animate there are more NP2 attachments when NP2 is concrete and NP1 is abstract. Secondly, when NP1 is inanimate it still attracts more RCs when it is concrete and modified by an abstract animate NP2.) This pattern was observed both in sentence production (corpus materials) and sentence reading (eye-tracking data), refuting previous suggestions of divergences between language production and language perception. These results are compatible with experience-based models of sentence parsing if they take into account structural and lexical frequencies.

Manuscript received December 2003

Revised manuscript received November 2004

PrEview proof published online August 2005

## REFERENCES

- Altmann, G. T. M. (2002). Learning and development in neural networks – the importance of prior experience. *Cognition*, 85, B43–B50.
- Baltazard D., & Kister, L. (1995, September). *Corrélation entre détermination et sélection d'un anaphorisé dans une structure N de N*. Paper presented at the Séminaire 'Anaphore et Référence'. Nancy (CRIN): 20–22 September.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 355–387.
- Bock, J. K., & Loebell, H. (1990). Framing sentences. *Cognition*, 35, 1–39.
- Bock, J. K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review*, 99, 150–171.
- Brybaert, M., & Mitchell, D. C. (1996). Modifier attachment in sentence processing: Evidence from Dutch. *Quarterly Journal of Experimental Psychology*, 49A, 664–695.
- Corley, M. M. B. (1996). *The role of statistics in human sentence processing*. Unpublished Ph.D. thesis, University of Exeter.
- Corrigan, R. (1986). The internal structure of English transitive sentences. *Memory and Cognition*, 14, 420–431.
- Crocker, M. W., & Brants, T. (2000). Wide-coverage probabilistic sentence processing. *Journal of Psycholinguistic Research*, 29, 647–669.
- Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition*, 30, 73–105.
- Cuetos, F., Mitchell, D. C., & Corley, M. M. B. (1996). Parsing in different languages. In M. Carreiras, J. E. Garcia-Albea, & N. Sabastian-Galles (Eds.), *Language processing in Spanish* (pp. 145–187). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- De Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology – Learning, Memory, and Cognition*, 18, 1001–1018.
- De Groot, A. M. B., & Hoeks, J. C. J. (1995). The development of bilingual memory: Evidence from word translation by trilinguals. *Language Learning*, 45, 683–724.
- Desmet, T., Brybaert, M., & De Baecke, C. (2002a). The correspondence between sentence production and corpus frequencies in modifier attachment. *Quarterly Journal of Experimental Psychology*, 55A, 879–896.
- Desmet, T., De Baecke, C., & Brybaert, M. (2002b). The influence of referential discourse context on modifier attachment in Dutch. *Memory and Cognition*, 30, 150–157.
- Desmet, T., & Gibson, E. (2003). Disambiguation preferences and corpus frequencies in noun phrase conjunction. *Journal of Memory and Language*, 49, 353–374.
- Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67, 547–619.
- Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Unpublished doctoral dissertation, University of Connecticut, Storrs, USA.
- Frazier, L., & Clifton, C. (1996). *Construal*. Boston: MIT Press.
- Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*, 32, 99–156.
- Gibson, E., Pearlmutter, N. J., & Torrens, V. (1999). Recency and lexical preferences in Spanish. *Memory and Cognition*, 27, 603–611.
- Gibson, E., & Schütze, C. T. (1999). Disambiguation preferences in noun phrase conjunction do not mirror corpus frequency. *Journal of Memory and Language*, 40, 263–279.
- Gibson, E., Schütze, C. T., & Salomon, A. (1996). The relationship between the frequency and the complexity of linguistic structure. *Journal of Psycholinguistic Research*, 25, 59–92.
- Grondelaers, S., Deygers, K., van Aken, H., van den Heede, V., & Speelman, D. (2000). Het ConDiv-corpus geschreven Nederlands. *Nederlandse Taalkunde*, 5, 356–363.

- Igoa, J. M., Carreiras, M., & Meseguer, E. (1998). A study on late closure in Spanish: Principle-grounded vs. frequency-based accounts of attachment preferences. *Quarterly Journal of Experimental Psychology*, *51A*, 561–592.
- Itagaki, N., & Prideaux, G. D. (1985). Nominal properties as determinants of subject selection. *Lingua*, *66*, 135–149.
- Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *20*, 137–194.
- Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 55–75). Oxford: Elsevier.
- MacDonald, J. L., Bock, J. K., & Kelly, M. H. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology*, *25*, 188–230.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.
- Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, *47*, 50–68.
- McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes*, *12*, 137–176.
- Meyer, A. S., & Bock, K. (1999). Representations and processes in the production of pronouns: Some perspectives from Dutch. *Journal of Memory and Language*, *41*, 281–301.
- Mitchell, D. C., & Brysbaert, M. (1998). Challenges to recent theories of crosslinguistic variation in parsing: Evidence from Dutch. In D. Hillert (Ed.), *Syntax and semantics: A crosslinguistic perspective* (pp. 313–335). San Diego, CA: Academic Press.
- Mitchell, D. C., Brysbaert, M., Grondelaers, S., & Swanepoel, P. (2000). Modifier attachment in Dutch: Testing aspects of Construal Theory. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 493–516). Oxford: Elsevier.
- Mitchell, D. C., & Cuetos, F. (1991). The origins of parsing strategies. In C. Smith (Ed.), *Current issues in natural language processing*. Austin: University of Texas, Centre for Cognitive Science.
- Mitchell, D. C., Cuetos, F., Corley, M. M. B., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research*, *24*, 469–488.
- Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.
- Ransdell, S. E., & Fischler, I. (1987). Memory in monolingual mode: When are bilinguals at a disadvantage? *Journal of Memory and Language*, *26*, 392–405.
- Rohde, D. L. T. (2002). *A connectionist model of sentence comprehension and production*. Unpublished Ph.D. thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.
- Schütze, C. T., & Gibson, E. (1999). Argumenthood and English prepositional phrase attachment. *Journal of Memory and Language*, *40*, 409–431.
- Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, *27*, 499–520.
- Spivey-Knowlton, M., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55*, 227–267.
- Sturt, P., Costa, F., Lombardo, V., & Frascioni, P. (2003). Learning first-pass structural attachment preferences with dynamic grammars and recursive neural networks. *Cognition*, *88*, 133–169.

- Tabor, W., Juliano, C., & Tanenhaus, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes*, *12*, 211–271.
- Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology – Learning, Memory, and Cognition*, *19*, 528–553.
- Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, *75*, 105–143.
- Wijnen, F. (1998, March). *Dutch Relative Clause Attachment in Two- and Three-site Contexts*. Poster presented at the 11th Annual CUNY Conference on Human Sentence Processing. New Brunswick, NJ, USA.
- Yamamoto, Y. (1999). *Animacy and reference*. Amsterdam: John Benjamins.

## Appendix A

These are the 32 items that were used in the eye-tracking experiment. The eight versions of each sentence were created by combining each of the four possible NP1s (between the first pair of parentheses) with both the plural and singular verb form (between the second pair of parentheses). The sentences of which the number is followed by a \* were those sentences that were coded as having an argument relation between the animate concrete NP1 and NP2 (in the extra analysis reported on page 470).

- 1\*. De oude bakker uit het dorp is ontzettend jaloers op de (prestaties / kastelen / ondernemingen / dochters) van de miljonair die indruk (maken / maakt) op de dorpelingen.
2. De gouverneur met de socialistische principes is verrast door de (vorming / auto / parochie / bisschop) van de priesters die (toont / tonen) dat de kerk in stilte toch ook evolueert.
3. De advocaat met de Siciliaanse voorouders was gefascineerd door de (fraude / wagen / firma / klant) van de chauffeurs die (aangeeft / aangeven) dat er veel geld circuleert in de transportwereld.
4. De directeur met de disciplinaire problemen bekritiseerde de (turnlessen / handboeken / leesgroepen / leerlingen) van de leraar die niet (voldoen / voldoet) aan de strenge eisen.
- 5\*. De professoren uit de politieke wetenschappen evalueren de (stijl / tekst / vakbond / chef) van de journalisten die erin (slaat / slagen) om de minister op zijn plaats te zetten.
6. De muzikleraar op het Gentse conservatorium bewonderde de (opleiding / apparatuur / fanfare / docent) van de muzikanten die erin (slaat / slagen) om jonge mensen aan te spreken.
- 7\*. De bestuursleden uit het Antwerpse haten de (strategieën / vlinderdassen / supporterclubs / sympathisanten) van de coach die hen (irriteren / irriteert) van bij het begin van het seizoen.
- 8\*. De jongens in de Rode duivels outfit bewonderen de (acties / tatoeages / trainingsclubs / liefjes) van de doelman die hen (verbazen / verbaast) omdat hij er aanvankelijk als een sukkel uitzag.
- 9\*. De parlementairen met de jarenlange ervaring discussieerden over de (visie / brief / partij / collega) van de politici die (illustreert / illustreren) waarom de Euro niet in Groot-Brittannië thuishoort.
- 10\*. De ministers uit Afghanistan hebben schrik van de (intenties / wapens / legers / vrienden) van de terrorist die hen (intimideert / intimideren) zodat er voorlopig nog geen maatregelen getroffen worden.
- 11\*. Het tienermeisje uit de Noorderkempen hoort over de (dromen / hoeden / fanclubs / supporters) van de zanger die haar (fascineren / fascineert) omdat ze totaal voorbijgestreefd zijn.
12. De critici uit de muziekwereld hadden het over de (optredens / handschoenen / orkesten / studenten) van de dirigent die iedereen (bekoren / bekoort) omdat ze zo flitsend zijn.
13. De rebellen uit de bezette gebieden vrezen de (orders / raketten / troepen / soldaten) van de generaal die (pogen / poogt) om de rebellen van de kaart te vegen.
14. De zakenman met de spectaculaire carrière spot met de (ideologie / woonwagen / generatie / geneesheer) van de hippies die (thuishoort / thuishoren) in de sixties.
15. De persploeg met de slechte reputatie schrijft over de (waanzin / speedboot / entourage / manager) van de filmsterren die (charmeert / charmeren) omwille van de flamboyante uitstraling.
16. De politieke analist uit Denemarken vertelde over de (invloed / biografie / commissie / adviseur) van de senatoren die (maakt / maken) dat het schandaal bekend geraakt bij het grote publiek.

- 17\*. De balletleraar met de jarenlange ervaring vertelt over de (danspassen / jurken / families / broers) van de ballerina die (bekoren / bekoort) door (hun / haar) elegantie.
18. De eigenzinnige column in het schoolkrantje vertelt over de (politiek / duikboot / regering / president) van de westerlingen die (probeert / proberen) om de Russen uit de Kaspische Zee te verdrijven.
- 19\*. De psycholoog met de vooruitstrevende visie introduceerde de (levenswijze / hobbykamer / vereniging / verzorger) van de bejaarden die (aantoont / aantonen) dat ouderen ook nog jong van geest kunnen zijn.
- 20\*. De agenten uit het elitekorps vrezen de (traditie / kroeg / clan / baas) van de maffialeden die hen (intimideert / intimideren) vanaf de jaren zeventig.
21. De bevolking zonder toekomstperspectieven respecteert de (beslissingen / documenten / organisaties / raadgevers) van de president die (garanderen / garandeert) dat er geen oorlog komt.
- 22\*. De student met de beloftevolle toekomst luisterde aandachtig naar de (toespraak / de computer / vereniging / assistent) van de lesgivers die hem (mededeelt / mededelen) dat hij de enige student is die de keuzevakken volgt.
- 23\*. De ijverige paters uit de trappistenabdij klaagden over de (campagne / drank / gilde / broer) van de brouwers die (maakt / maken) dat het trappistenbier minder goed verkoopt.
24. De deelnemers aan de Ronde van Frankrijk kennen de (fratsen / truitjes / clubs / sponsors) van de wielrenner die hen (vervelen / verveelt) omdat ze de wielersport telkens opnieuw belachelijk maken.
- 25\*. De cafébazin uit de arme volkswijk wantrouwt de (beloftes / producten / bedrijfjes / collega's) van de handelaar die (garanderen / garandeert) dat haar leven zal veranderen.
26. De onderzoeker met de vernieuwende ideeën is geïnteresseerd in de (resultaten / artikels / onderzoeksgroepen / studenten) van de professor die (breken / breekt) met de klassieke theorie.
27. De literatuurdeskundigen op de boekenbeurs bewonderen de (stelling / verhandeling / uitgeverij / promotor) van de schrijvers die (probeert / proberen) om de mensen wakker te schudden.
- 28\*. De koning met de dictatoriale trekjes bekritiseert de (uitspraken / rapporten / comités / medewerkers) van de gouverneur die (onderstrepen / onderstreept) dat het land slecht bestuurd wordt.
29. De parlementsleden bij de Europese Unie zijn ontevreden over de (initiatieven / documenten / agentschappen / secretaresses) van de commissaris die (doen / doet) geloven dat de parlementairen te hoge lonen krijgen.
- 30\*. De volksvertegenwoordigers in het Vlaamse Parlement praten over de (verklaring / brief / coalitie / bondgenoot) van de politici die (reageert / reageren) tegen de invoering van het migrantenstemrecht.
31. De legerleiding in Jeruzalem heeft zich schrik laten aanjagen door de (aanvalsplannen / atoombommen / verzetsbewegingen / paracommando's) van de rebellenleider die (aangeven / aangeeft) dat het deze keer wel tot een serieuze confrontatie kan komen.
- 32\*. De bouwvakkers uit de Vlaamse Ardennen lachen met de (voorspellingen / geschriften / sektes / volgelingen) van de goeroe die het (hebben / heeft) over een nieuwe wereld zonder oorlogen en armoede.

## Appendix B

This appendix contains the ANOVAs that were performed on the regions prior to the disambiguating regions. We performed ANOVAs with two repeated measures: head type of NP1 and attachment site. This was done both over participants ( $F_1$ ) and over items ( $F_2$ ).

### CRRT

Region 1 (the beginning of the sentence) was identical in all conditions, and as expected none of the effects was statistically significant (main effect of head type:  $F_1$  and  $F_2 < 1$ ; main effect of attachment site:  $F_1$  and  $F_2 < 1$ ; interaction:  $F_1(3,138) = 2.62$ ;  $p = .05$ ;  $F_2(3,90) = 1.70$ ,  $p = .17$ ).

The ANOVA on Region 2 (the first attachment site NP1) revealed a main effect of NP1 type in the analysis over participants,  $F_1(3,138) = 3.64$ ,  $p < .05$ . The effect was only marginal in the analysis over sentences,  $F_2(3,90) = 2.37$ ,  $p = .08$ . Planned comparisons showed that this effect was due to the fact that abstract nouns were read more slowly than the concrete nouns,  $F_1(1,46) = 10.26$ ,  $p < .01$ ;  $F_2(1,30) = 8.91$ ,  $p < .01$ . A series of t-tests indicated that this effect was not due to differences in length or frequency between the different nouns that were used in the region (all  $t < 1.68$ , all  $p > .10$ ). There was no main effect of attachment site,  $F_1(1,46) = 1.39$ ,  $p = .24$ ;  $F_2(1,30) = 1.12$ ,  $p = .30$ , nor a significant interaction,  $F_1(3,138) = 1.37$ ,  $p = .26$ ;  $F_2(3,90) = 1.99$ ,  $p = .12$ .

The words in the third region (the PP made up of the preposition 'van' and the second attachment site NP2) were the same in all eight versions of a stimulus set. Yet, a significant main effect of NP1 type was found,  $F_1(3,138) = 10.04$ ,  $p < .001$ ;  $F_2(3,90) = 8.50$ ,  $p < .001$ . Post-hoc tests (Tukey) indicated the following order as a function of NP1 type: inanimate-abstract < inanimate-concrete = animate-abstract < animate-concrete (although only the difference between inanimate-abstract and animate-concrete exceeded the .05 significance level both in the analysis over participants and over items). That is, the modification of NP1 by NP2 (which in all our stimuli referred to people) was hardest when NP1 already referred to a concrete person, and easiest when NP1 referred to an abstract notion (such as 'performance', 'education', 'fraud', 'style'). There was no main effect of attachment site,  $F_1(1,46) = 2.29$ ,  $p = .14$ ;  $F_2(1,30) = 1.23$ ,  $p = .28$ , and no interaction ( $F_1$  and  $F_2 < 1$ ).

At Region 4 (the relative pronoun 'die') there was no main effect of NP1 type,  $F_1(3,138) = 1.94$ ;  $p = .14$ ;  $F_2(3,90) = 2.17$ ,  $p = .10$ , no main effect of attachment site,  $F_1(1,46) = 1.47$ ,  $p = .23$ ;  $F_2(1,30) = 1.13$ ,  $p = .30$ , and no interaction between both variables,  $F_1(3,138) = 1.42$ ,  $p = .24$ ;  $F_2 < 1$ .

### FPRT

In Region 1, the first-pass reading times are identical to the cumulative region reading times, because there are no earlier regions that participants could go back to while reading this region. Consequently, there was no need to reanalyse this region.

In Region 2, FPRTs were shorter for concrete NP1s than for abstract NP1s, in line with the pattern we observed in the CRRTs. The omnibus ANOVA revealed a significant effect of NP1 type,  $F_1(3,138) = 4.79$ ,  $p < .01$ ;  $F_2(3,90) = 3.12$ ,  $p < .05$ , and a contrast of concrete NP1s vs. abstract NP1s confirmed that this distinction was the origin of the effect in the omnibus analysis,  $F_1(1,46) = 9.88$ ,  $p < .01$ ;  $F_2(1,30) = 8.97$ ,  $p < .01$ . There was no significant effect of attachment site ( $F_1$  and  $F_2 < 1$ ) and no significant interaction,  $F_1 < 1$ ;  $F_2(3,90) = 1.13$ ,  $p = .34$ .

In Regions 3 and 4, the FPRTs were also very much in line with the CRRTs. They were shortest for inanimate-abstract NP1s and longest for animate-concrete NP1s. This effect of

NP1 type was significant in the analysis over participants (Region 3:  $F_1(3,138) = 2.87, p < .05$ ; Region 4:  $F_1(3,138) = 2.75, p < .05$ ) and marginally significant in the analysis over items (Region 3:  $F_2(3,90) = 2.64, p = .05$ ; Region 4:  $F_2(3,90) = 2.29, p = .08$ ). There was no significant effect of attachment site (Region 3:  $F_1$  and  $F_2 < 1$ ; Region 4:  $F_1(1,46) = 3.88, p = .06$ ;  $F_2(1,30) = 2.65, p = .11$ ) and no significant interaction (Region 3:  $F_1$  and  $F_2 < 1$ ; Region 4:  $F_1(3,138) = 1.29, p = .28$ ;  $F_2(3,90) = 1.02, p = .39$ ).

### Regressions

The percentage of regressions of Region 1 was zero in all conditions because there are no earlier regions to go back to. Because Region 4 was very short (the relative pronoun 'die') there were too few observations to perform analyses on this region.

Very few regressions were made from Region 2 (NP1 type) and there was no effect of NP1 type on this variable ( $F_1$  and  $F_2 < 1$ ), despite the fact that abstract words took longer to read than concrete words (see the analyses of CRRT and FPRT). There was also no significant main effect of attachment site ( $F_1$  and  $F_2 < 1$ ) and no significant interaction,  $F_1(3,138) = 2.64, p = .06$ ;  $F_2(3,90) = 1.95, p = .13$ .

At Region 3, the significant main effect of NP1 type in CRRTs and FPRTs was also reflected in the percentage of regressions,  $F_1(3,138) = 6.36, p = .001$ ;  $F_2(3,90) = 5.98, p = .01$ . There was no significant main effect of attachment site nor a significant interaction (all  $F < 1.02$ ).