

UMOP 38

UMOP 38 · 2011

Processing Linguistic Structure

Edited by
Jesse A. Harris
and Margaret Grant

UMOP 38: Processing Linguistic Structure

Jesse A. Harris and Margaret Grant (eds.)

© Copyright 2011 by Jesse A. Harris and Margaret Grant
All Rights Reserved

Reproduced and distributed by

GLSA
(Graduate Linguistic Student Association)
Department of Linguistics
226 South College
University of Massachusetts, Amherst
Amherst, MA 01003
U.S.A.

glsa@linguist.umass.edu

ISBN: 1-4663-6907-8

EAN-13: 9781466369078

Table of Contents

<i>On the Grammatical Status of PP-Pied-Piping in English: Results from Sentence-Rating Experiments</i>	1
Seth Cable and Jesse A. Harris	
<i>Dynamic interpretation: Finding an antecedent for VPE</i>	23
Lyn Frazier and Charles Clifton, Jr.	
<i>D-linking and Memory Retrieval: The Annoying Case of Sluicing</i>	37
Lyn Frazier and Charles Clifton, Jr.	
<i>From nounphrase ellipsis to verbphrase ellipsis: The acquisition path from context to abstract reconstruction</i>	53
Tilbe Goksun, Tom Roeper, Kathy Hirsh-Pasek, and Roberta M. Golinkoff	
<i>Discourse Structure and Syntactic Parallelism in VP Ellipsis</i>	75
Christina Kim and Jeffrey Runner	
<i>Processing Null-Operator Structures</i>	103
Jason Overfelt	
<i>Processing effects of an indeterminate future: Evidence from self-paced reading</i>	131
Adrian Staub and Charles Clifton, Jr.	
<i>On-line costs for predicting upcoming syntactic structure</i>	141
Tessa Warren and Michael Walsh Dickey	

On the Grammatical Status of PP-Pied-Piping in English: Results from Sentence-Rating Experiments*

Seth Cable and Jesse A. Harris

University of Massachusetts Amherst

1. Introduction

We report a series of sentence-rating experiments that were designed to test recent claims about the grammaticality of sentences formed with certain movement operations related to question formation in matrix and subordinate clauses in English. Specifically, we tested the claim that so-called ‘pied piping’ of prepositional phrases in English is largely restricted to main clauses, and is rather marginal in subordinate clauses.

In English question formation, so-called *wh*-elements like *who*, *what*, and *how* are commonly analyzed as having moved to a position other than that which they were subcategorized for. For example, the question word *who* in (1b) is commonly viewed as having originated in object position t_1 , the position occupied by John in (1a). Its surface appearance at the left periphery of the clause is attributed to a movement operation that displaces the *wh*-word into its observed position.

(1) Simple Wh-Movement in English

- a. She left John
- b. Who₁ did she leave t_1 ?

*For Experiment 1, we wish to thank Margaret Grant for her assistance in creating items, as well as, collecting data, for which we also thank Aynat Rubinstein. For Experiment 2, we extend our thanks to Lyn Frazier for discussing our materials and design with us, to Kyle Johnson and Peggy Speas for allowing us to recruit participants from their courses, to John J. McCarthy for allocating the funds to pay subjects on Amazon’s Mechanical Turk, and to Alex Drummond for developing the Ibex farm presentation software and his help implementing our experiment. We are grateful to Chuck Clifton and Brian Dillon for comments on an earlier draft. All errors are naturally our own.

In addition to the basic case in (1), in which only a *wh*-element is subject to movement, the moved *wh*-element can co-occur with another phrasal element, making for a more complex moved constituent. Such cases, illustrated below, are commonly referred to as ‘pied-piping’ (Ross 1967), a term that metaphorically suggests the *wh*-word has lead the other phrasal material along as it moves to the left periphery.

(2) **Pied-Piping in English**

- a. She left that guy.
- b. [Which guy]₁ did she leave *t*₁ ?

Interestingly, when a *wh*-element is complement to a preposition, there appears to be some optionality in whether the preposition is ‘pied-piped’ when the *wh*-element undergoes movement. That is, it would appear that English freely permits both the structures in (3). In sentence (3a), the *wh*-word has not pied-piped the higher PP. Such structures are commonly referred to as ‘preposition stranding’ or ‘P-stranding’. In sentence (3b), the *wh*-word has pied-piped the PP, a structure referred to as ‘PP-pied-piping’.

(3) **The Optionality of Pied-Piping PP in English**

- a. Who₁ did she leave [_{PP} with *t*₁]?
- b. [_{PP} With whom]₁ did she leave *t*₁ ?

Although both the structures in (3) are commonly reported in the syntactic literature as ‘acceptable’, the grammatical status of PP-pied-piping structures is somewhat unclear, especially when compared to their preposition stranding counterparts. While speakers recognize structures like (3b) as English, such structures are not particularly colloquial. It is sometimes said that such structures are limited to particular registers, but often what is meant by ‘register’ in this context is unclear. After all, structures like (3b) are no longer a regular occurrence in formal written English either.

Perhaps not coincidentally, English PP-pied-piping poses analytic difficulties for current theories of pied-piping. Both the theories of Heck (2008) and Cable (2010) are at first glance challenged by the seemingly free alternation between (3a) and (3b). As noted by both Heck (2008) and

Cable (2010), these two theories predict that (3a) should serve to ‘block’ (3b).¹

In response to this challenge, both Heck (2008) and Cable (2010) argue that PP-pied-piping in English possesses a special status, that of ‘massive pied-piping’. In brief, the term ‘massive pied-piping’ refers to the ability for the locality conditions on pied-piping to be marginally violated in non-embedded (matrix) clauses. For example, English typically does not allow a *wh*-word to pied-pipe a noun phrase (NP) it is contained within (4). This condition can be understood as a locality condition on the pied-piping relation itself: a lexical category like NP cannot intervene between the *wh*-word and the edge of the pied-piped phrase (Heck 2008; Cable 2010). Structures that violate this constraint contain an NP in the pied-piped phrase, leading to the impression that the size of the pied-piping is larger than usual, or otherwise ‘massive’. For example, in the ill-formed (4b.i), the NP *a picture* in the pied-piped phrase *a picture of whom* intervenes between the left phrase boundary and the *wh*-word *whom*, whereas movement of *who* to a higher clausal position in (4b.ii) appears to obviate the restriction.

(4) **Massive Pied-Piping of NPs in English**

a. *In Matrix Clause*

- i. ?? [A picture of whom] did Bill take at the party?
- ii. Who did Bill take [a picture of] at the party?

b. *In Embedded Clause*

- i. * I wonder [a picture of whom] Bill took at the party.
- ii. I wonder who Bill took [a picture of] at the party.

As noted above, a defining feature of massive pied-piping is that it is restricted to non-embedded clauses. Put differently, the locality constraints governing pied-piping structures are (in some cases) weaker in main clauses than they are in subordinate clauses. Although Heck (2008) and Cable (2010) explain these contrasts differently, both share the view that embedded clauses reveal the true, general conditions on pied-piping in a given language. The seemingly broader range of pied-piping structures in main

¹ Heck (2008) and Cable (2010) differ in the exact way in which this alleged ‘blocking’ occurs. The details here would take us too far afield in the present work. The interested reader is referred to Heck (2008: 117–127) and Cable (2010: 166–173).

clauses is merely due to independent factors that can mitigate the severity of the grammatical violations in question.²

With this in mind, both Heck (2008) and Cable (2010) note the relative ill-formedness of English PP-pied-piping in subordinate clauses.³

(5) **Pied-Piping of PPs in English Limited to Main Clauses**

- a. (?) [With whom] did she leave?
- b. * I wonder [with whom] she left.

This contrast suggests that PP-pied-piping structures in English should be categorized as massive pied-piping structures. Note that a similar contrast does not seem to exist for ‘determiner pied-piping’ structures (DP-pied-piping) like those in (6), structures that all theories of pied-piping find unproblematic.

(6) **Pied-Piping by Wh-Determiners in English is Not Massive Pied-Piping**

- a. [Which book] did she borrow?
- b. I wonder [which book] she borrowed.

The difference between the patterns in (5) and (6) reinforces the notion that pied-piping of PPs has a distinct grammatical status in English, one that separates it from pied-piping by determiners and possessors.⁴

Given these facts, Heck (2008) and Cable (2010) indeed conclude that PP-pied-piping in English is a case of ‘massive pied-piping’. Thus,

² According to Heck (2008: 297–315), these ‘independent factors’ are the possibility of feature movement from the *wh*-word in main clauses. According to Cable (2010: 190–198), they are the ability for Q-particles in main clauses not to bear the [*wh*]-features that otherwise force Agreement with *wh*-words.

³ As noted by (Heck 2008: 123), this contrast has been observed by many other scholars. Heck cites at least eight other works noting this contrast, the first of which appears to be Allen (1980).

⁴ Citing prior literature, Heck (2008) notes that pied-piping of PPs in subordinate clauses greatly improves in sentences where P-stranding is not permitted. As illustrated below, such cases of ‘forced pied-piping’ do seem to be licit for PPs in subordinate clauses.

- (i) I wonder [[in what sense] he was responsible]
- (ii) * I wonder [[what sense] he was responsible in]

In this paper, we put aside such cases of ‘forced pied-piping’, restricting our attention to cases where PP-pied-piping appears to optionally vary with P-stranding. That is, all our items in Experiment 1 were instances of optional (non-forced) pied-piping; see Appendix A for details.

as revealed in embedded clauses (5b), such structures do indeed violate the locality conditions on English pied-piping; this violation is simply mitigated in main clauses like (5a).⁵ Consequently, the analytic problem raised by the seemingly free variation between (3a) and (3b) disappears, as only (3a) is entirely licit.

While this line of analysis rescues Heck (2008) and Cable (2010) from the challenge posed by the data in (3), one is right to criticize its empirical basis. The key notion – that pied-piping of PP is ‘massive pied-piping’ in English – is supported by only a few crucial example sentences.⁶ Further complicating things is the fact that the key comparison between (5) and (6) is not truly a minimal one. Since one is comparing PP-pied-piping to determiner pied-piping, various extraneous changes in lexical choice and syntactic structure separate the sentences to be compared. Consequently, one might rightly worry whether the putative contrast between (5) and (6) is simply an artifact of the particular examples chosen.

For these reasons, we sought to more rigorously test the proposal that pied-piping of PPs in English is ‘massive pied-piping’. We developed a series of sentence-rating experiments to test the following two predictions, which follow from accounts like Heck (2008) and Cable (2010), albeit for different reasons.

(7) **Experimental Predictions to be Tested**

- a. Pied-piping of PPs is less acceptable in English subordinate clauses than it is in main clauses.
- b. The contrast between pied-piping in main clauses vs. subordinate clauses is seen only for PP pied-piping. A comparable contrast is not found for pied-piping of DPs by *wh*-determiners.

The following sections describe the structure of these experiments in more detail. In Section 2, we review our first experiment, which supported pre-

⁵ Of course, it must be noted that speakers find PP-pied-piping in main clauses is much better than other cases of massive pied-piping, such as that in (4a.i). Neither Heck (2008) nor Cable (2010) account for this. However, one explanation may be that speakers find PP-pied-piping more ‘familiar’, through their exposure to literature from earlier stages of the English language where PP-pied-piping was more prevalent. Also see the discussion in Heck (2008: 123–124).

⁶ Again, however, the key empirical observation has been informally corroborated by several scholars, as cited by Heck (2008: 123).

diction (7a). Section 3 presents the second experiment, which was found to support prediction (7b).

2. Experiment 1

Prediction (7a) states that pied-piping of PPs is less acceptable in English subordinate clauses than in main clauses. That is, if pied-piping of PPs in English is truly ‘massive pied-piping’, then we should find the judgment paradigm in (8) can be observed for PP-pied-piping. Specifically, we should find a greater contrast in the acceptability ratings between the embedded pairs displayed in (8b.i-ii) than their matrix counterparts displayed in (8a.i-ii). In other words, the effect of PP-pied-piping should be greater for embedded clauses than for matrix clauses.

(8) **Judgment Pattern, If Pied-Piping of PPs is Massive Pied-Piping**

- a. In Matrix Clause
 - i. ?? [With whom] did he dance?
 - ii. Who did he dance with?
- b. In Embedded Clause
 - i. * I wonder [with whom] he danced.
 - ii. I wonder who he danced with.

Our first experiment sought to test this prediction. The details of its design and results are reported below.

2.1 Design and materials

Experiment 1 was developed as a 2x2 factorial design, in which we crossed Clause type (Matrix or Subordinate clause) with Dependency type (PP-pied-piping or Preposition stranding). In (9) below, we illustrate the paradigm with an example quartet from the experiment:

- (9) a. With whom did he dance? (Matrix: PP-pied-piping)
- b. I wonder with whom he danced. (Subordinate: PP-pied-piping)
- c. Who did he dance with? (Matrix: Prep stranding)
- d. I wonder who he danced with. (Subordinate: Prep stranding)

All of the Pied-piping cases contained the prescriptive relative pronoun *whom*, a potential confound that was eliminated in Experiment 2. The

materials consisted of 8 quartets following the pattern in (9); see Appendix A for a complete list of items.

2.2 Participants and procedures

Forty undergraduates from the University of Massachusetts Amherst participated in the study for course credit. The experiment took less than 1/2 hour on average to complete. Participants were invited to take a break between trials at any point during the experiment, and were given a planned break halfway through the experiment to prevent fatigue. The items were presented in randomized counterbalanced order with items from 5 sub-experiments, unrelated to the manipulation presented here, and 10 non-experimental filler sentences, for a total of 102 experimental trials per session. Items were shown in a Latin Square design so that each participant saw one and only one condition from each experimental quartet.

Participants were instructed to read each sentence carefully and then rate its acceptability on a 5-point Likert scale. Each sentence item and the acceptability scale, as in (10), were simultaneously presented on a computer screen with the program Linger,⁷ which recorded responses and reaction times. Prior to the experiment, participants were instructed verbally and then again in writing with a guided practice. Participants were invited to ask the experimenter for assistance at any point during the experiment.

- (10) a. I wonder with whom he danced?
b. How acceptable was that sentence?
i. Terrible
ii. Pretty bad
iii. So so
iv. Pretty good
v. Perfect

2.3 Results

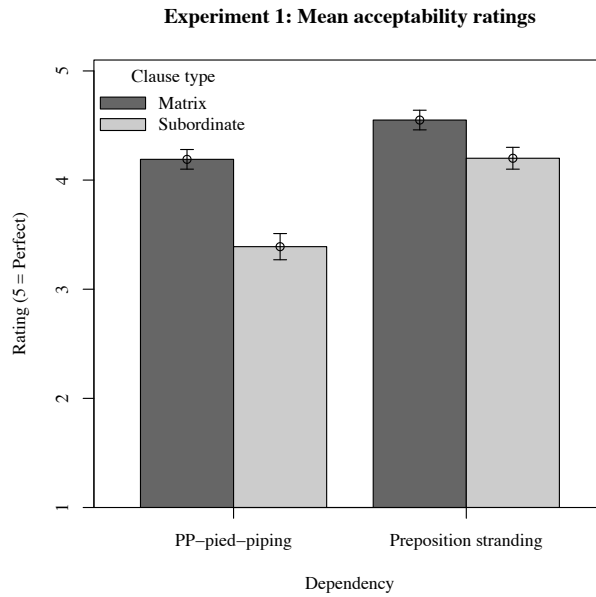
Prior to analysis, the data were cleaned of outliers on the basis of reaction time. Scores with reaction times registering over 3 standard deviations were removed from the data. Removing these outliers resulted in less than 3%

⁷ Linger is available at <http://tedlab.mit.edu/~dr/Linger/>

data loss for any one condition, and less than 2% data loss for the entire data set. All data analysis was computed using the R statistical language suite R Development Core Team (2008).

2.3.1 Acceptability judgments

We first present the mean acceptability ratings and standard errors in Figure 1.



	PP-pied-piping	Preposition stranding	Mean
Matrix	4.19 (0.09)	4.55 (0.09)	4.37 (0.07)
Subordinate	3.39 (0.12)	4.20 (0.10)	3.80 (0.08)
Mean	3.79 (0.08)	4.38 (0.07)	

Figure 1: Mean acceptability judgment ratings (standard errors in parentheses) by condition for Experiment 1.

We used a linear mixed-effects regression model to analyze the data, treating subjects and items as simple random effects (for discussion of this method in the context of psycholinguistic designs see, e.g., Baayen, 2008;

Baayen et al., 2008). In our model, we crossed the planned factors of Clause type and Dependency, as illustrated in (11) below; any statistical results we report were obtained from this model, using the lmer package (Bates & Maechler 2009) to compute significance values. All significant results are reported.

$$(11) \quad \text{Rating} \sim \text{Clause} \times \text{Dependency} + (1|\text{Subject}) + (1|\text{Item})$$

First, as expected, there was a significant main effect of Clause type: items with subordinate clauses were rated significantly lower ($M = 3.80$, $SE = 0.08$) than their matrix counterparts ($M = 4.37$, $SE = 0.07$), $t = -6.37$, $p < 0.001$. This cost for subordinate clauses is not surprising; as subordinate clauses are more complex, they should be more difficult to process, and hence be judged as less acceptable. Second, we observed a main effect of Dependency: items with Preposition stranding ($M = 4.38$, $SE = 0.07$) were rated as significantly more acceptable than items with PP-pied-piping ($M = 3.79$, $SE = 0.08$), $t = 2.75$, $p < 0.01$. Again, this effect is not particularly surprising, as we expect that the PP-pied-piping construction is probably far less frequent than constructions with preposition stranding in modern English.

Crucially, however, we observed a significant interaction between Clause type and Dependency, such that Clause type affected the acceptability rating of PP-pied-piping structures more than it affected those of Preposition stranding structures. In particular, the extent to which a subordinate clause lowered acceptability ratings as compared to matrix counterparts was significantly greater for PP-pied-piping structures ($d = 0.8$) than it was for Preposition stranding structures ($d = 0.35$), $t = 2.65$, $p < 0.01$. The direction of the interaction is particularly clear in Figure 1. To determine whether this effect was spurious, we removed the interaction term from the model and fit the data to this simpler model. Again, we observed an effect of Clause type and an effect of Dependency. However, comparing the two models revealed that removing the interaction term did not result in a better fitting model. Indeed, the first model, which included the crucial interaction term, was a significantly better fit, $\chi^2 = 6.99$, $p < 0.001$, even though it was a more complex model.

Finally, although we did collect reaction time data from the experiment, we do not present it in detail here. We choose to omit the data,

because it is of only marginal interest. Recall that our reaction times include the time spent reading the sentence, as well as the time spent rating the sentence, as subjects viewed both the target sentence and the rating scale simultaneously (10). As the subordinate condition was longer across the board, we would expect that conditions with subordinate clauses would elicit longer reading times – regardless of dependency type – for relatively uninteresting reasons. That is indeed the only effect regarding reaction times that we found in this experiment.

2.4 Discussion

As noted above, the results of Experiment 1 support the predictions in (7a); there was a significant difference between the effect of PP-pied-piping in main clauses and subordinate clauses. Furthermore, as the difference between matrix and subordinate structures was significantly greater for PP-pied-piping than for Preposition stranding, the effect cannot be reduced to a contrast between matrix and subordinate clauses. Thus, Experiment 1 confirms that the key judgment pattern in (8) does hold for PP-pied-piping, which supports the claim that such pied-piping should be categorized as ‘massive pied-piping’.

However, since only PP-pied-piping was examined in Experiment 1, there remains the question of whether the effects observed are specific to that construction. After all, if all pied-piping structures exhibited these effects, then the conclusion that PP-pied-piping deserves special categorization is defeated. For this reason, we sought in Experiment 2 to examine the possible existence of these effects in ‘non-massive’ pied-piping.

3. Experiment 2

Prediction (7b) states that the effect of clause type observed for PP-pied-piping should not be found for instances of ‘non-massive’ pied-piping, such as pied-piping by *wh*-determiners. That is, if the effect observed in Experiment 1 is truly due to PP-pied-piping having a special status within pied-piping constructions, then the judgment paradigm in (8) should not be observed for those other constructions.

Unfortunately, due to independent features of English grammar, one cannot make a perfect comparison between the pied-piping of PPs and ‘non-massive’ pied-piping. The problem is that all cases of ‘non-massive’ pied-

piping in English are also not optional. For example, while both (3a) and (3b) are possible in English, only (12b) is grammatically licit.

(12) **Non-Massive Pied-Piping (in English) is Not Optional**

- a. * Which did she write book.
- b. [Which book] did she write?

Thus, one cannot perfectly adapt Experiment 1 for, e.g., pied-piping by *wh*-determiners. Instead, the best one can do is to compare the interaction between clause-type and pied-piping. That is, if the effect of clause type is greater for PP-pied-piping than for DP-pied-piping, then this lends support to prediction (7b), and thus the special status of PP-pied-piping in English, in support of analyses like Heck (2008) and Cable (2010). Of course, such an effect would also be open to alternative analyses, which will be discussed in Section 3.4 below. Before we come to that, however, we lay out the experimental design and results.

3.1 Design and materials

Experiment 2 was also developed as a 2x2 factorial design, crossing Clause type (Matrix or Subordinate) with Pied-piping type (DP-pied-piping or PP-pied-piping), as illustrated in (13). There were 12 such quartets in all.

- (13) a. Which book did she write? (Matrix: DP-pied-piping)
- b. It's obvious which book she wrote. (Subord: DP-pied-piping)
- c. About what did she write? (Matrix: PP-pied-piping)
- d. It's obvious about what she wrote. (Subord: PP-pied-piping)

The complete materials may be found in Appendix B.

3.2 Participants and procedures

We conducted our experiment twice. In the first instance, 91 members from the general population participated in Experiment 2; they were recruited on Amazon's Mechanical Turk, an internet-based service where workers anonymously answer short questions or perform small tasks in exchange for payment. Workers who elected to participate in our experiment were given \$5 as compensation. Only workers who had performed at least 50 previous questions and received a 98% approval rating or above were permitted to

participate in the experiment. Prior to taking the questionnaire, participants were asked to fill out a brief survey on their linguistic history. Self reported non-native English speakers were excluded from the analysis. In the second instance, we sought to replicate our results with undergraduates from the University of Massachusetts Amherst.

Since we did not have direct contact with our participants, we enacted several stringent controls to reduce the possibility of deceptive or disengaged participants. First, we recorded an MD5 hash of the IP address of all our participants to reduce the possibility that participants were repeating the experiment. If identical IP addresses were recorded, all of that participant's data were removed. Second, immediately after reading the instructions, participants were asked to answer three comprehension questions from Harris & Potts (to appear), designed to be particularly difficult for non-native speakers. Incorrectly answering any one of these questions was grounds for removal. Third, we included 4 undeniably ungrammatical or non-sensical sentences with failures of number, case agreement, or sensicality (see Appendix B for these items). Any participant who rated any one of these items greater than a 3 was excluded from analysis. Twenty-five subjects were eliminated from the data on this basis. Fourth, we looked at the responses over the course of the experiment. Participants who showed little to no variation in their responses – either in their ratings or the speed by which they responded – were excluded on the presumption that they were not genuinely performing the task. In order to create a balanced dataset, we excluded 18 participants from the experiment.⁸ Finally, we removed observations with reaction times that were greater than 3 standard deviations from the mean, which, as before, resulted in less than 3% data loss for any condition, and less than 2% data loss across the entire experiment.

After a short, guided practice, participants rated the items for acceptability on a 7-point Likert scale as in (14) below, in which 1 was designated as “Completely unacceptable” and 7 as “Completely acceptable.”

- (14) Which book did she write?
 How acceptable is this sentence? (Please use entire scale.)
 (Completely unacceptable) 1 2 3 4 5 6 7 (Completely acceptable)

Items were presented in a Latin Square design, counterbalanced and

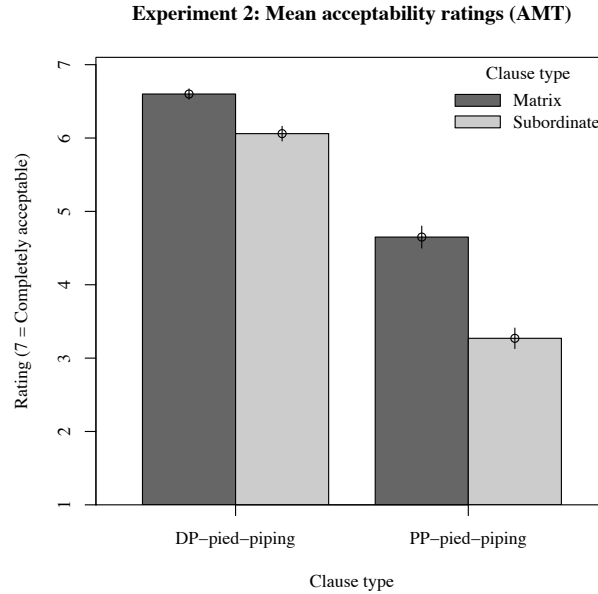
⁸ Removing these subjects did not affect the general patterns observed in the data.

randomly interspersed with 48 items from another experiment on Pied-piping, 61 genuine filler items, and 4 catch sentences, as discussed above, for a total of 125 items per session. Ratings and reaction times were presented and recorded over the Internet using Ibx farm. Participants from the first group were told that they had 1 hour and 20 minutes to complete the experiment, which took approximately 1/2 hour on average. Participants were invited to complete the questionnaire at their own pace, and were invited to take breaks as needed; after rating each item, participants saw a screen with the following instructions “When you’re ready, press a key to see the next sentence.”

3.3 Results

We first present the results from the group who were recruited from Amazon’s Mechanical Turk (the AMT group), and then, as the results are virtually identical, briefly report the results obtained from student participants at the University of Massachusetts Amherst (the UMASS group). Means and standard errors from the AMT group are provided in Figure 2 below. The procedure for data analysis was identical to that described in Experiment 1, modulo the different planned factors. We again observed a cost for subordinate clauses: pied-piping in matrix clauses ($M = 5.63$, $SE = 0.10$) was rated significantly better than pied piping in subordinate clauses ($M = 4.67$, $SE = 0.12$), $t = -3.54$, $p < 0.001$. In addition, there was a main effect of pied-piping type, such that participants rated pied-piping of PPs ($M = 3.96$, $SE = 0.11$) significantly worse than pied-piping of DPs ($M = 6.33$, $SE = 0.06$), $t = -13.22$, $p < 0.001$. Finally, we observed an interaction between the pied-piped element and embedding: the negative effect of the subordinate clause was greater when the pied-piped element was a PP ($d = 1.38$) than when it was a DP ($d = 0.54$), $t = -4.03$, $p < 0.001$. As the interaction is the crucial effect, we again wanted to test whether we were justified in retaining it in our model. Thus, as before, we removed the interaction and fit the data to a simpler model with just Clause type and Pied-piping element as additive factors. Although more complex, the model with the interaction term was a significantly better fit, $\chi^2 = 16.07$, $p < 0.001$, justifying its explanatory role in the model.

We also collected the times that subjects spent on each trial. As before, this measure includes both the time to read the item and to rate it; we

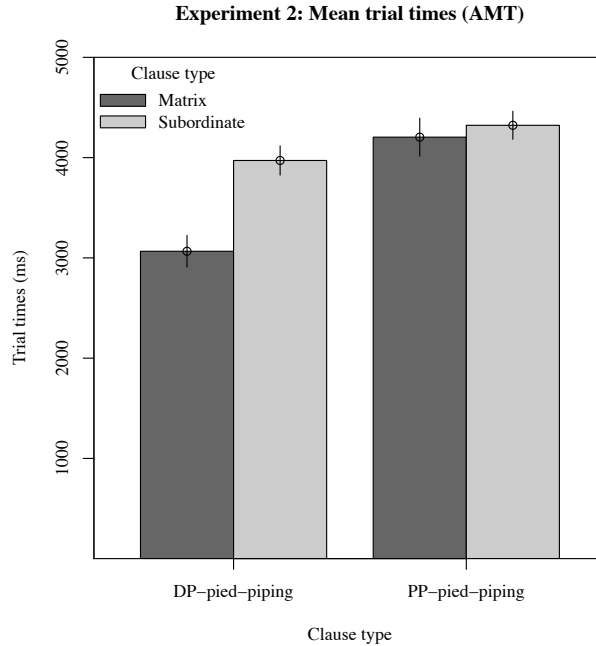


	DP-pied-piping	PP-pied-piping	Mean
Matrix	6.60 (0.07)	4.65 (0.15)	5.63 (0.10)
Subordinate	6.06 (0.10)	3.27 (0.14)	4.67 (0.12)
Mean	6.33 (0.06)	3.96 (0.11)	

Figure 2: Mean acceptability judgment ratings (standard errors in parentheses) by condition for the AMT group in Experiment 2.

call this measure ‘trial time’ to reflect the fact that it is distinct from traditional reading time or decision time measurements. An additional caveat is in order; we hesitate to interpret the data too closely for two reasons. First, participants were not monitored during the experiment, and thus trial times may include extraneous actions unrelated to the experiment. Second, we do not know enough about the method used to record the trial time measurement – in particular, the extent to which collection is delayed or modified during transmission across the Internet to the Ibex server. Nevertheless, we believe the results to be intriguing, and opt to present them here in Figure 3, cautions and concerns notwithstanding.

We observed a main effect of Clause type, such that items with sub-



	DP-pied-piping	PP-pied-piping	Mean
Matrix	3065.98 (158.63)	4204.21 (189.47)	3631.08 (127.75)
Subordinate	3971.61 (146.00)	4322.39 (140.19)	4147.00 (101.56)
Mean	3517.20 (110.94)	4263.51 (117.57)	

Figure 3: Mean trial times in milliseconds (standard errors in parentheses) by condition for the AMT group in Experiment 2.

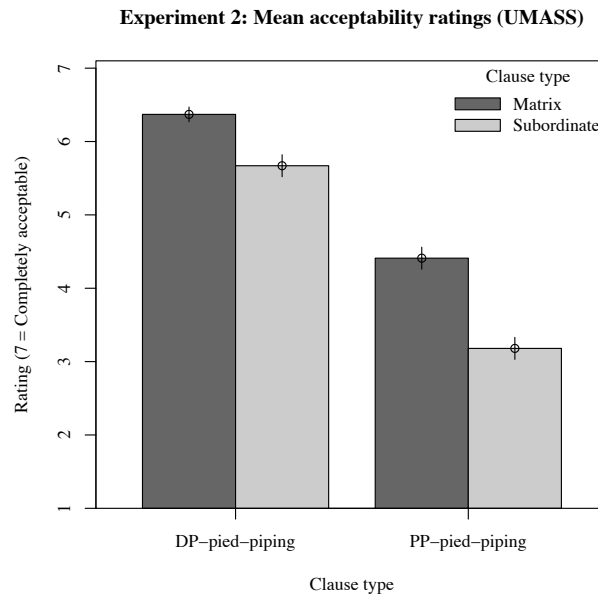
ordinate clauses elicited significantly slower trial times ($M = 4147\text{ms}$, $SE = 102$), than Matrix counterparts ($M = 3631\text{ms}$, $SE = 128$), $t = 4.435$, $p < 0.001$. In addition, there was a main effect of pied-piping element: PP pied-piping items elicited significantly slower trial times ($M = 4263\text{ms}$, $SE = 117.57$), than DP pied-piping counterparts ($M = 3517\text{ms}$, $SE = 110.94$), $t = 5.15$, $p < 0.001$. Finally, the clause type differentially affected the speed with which participants completed a trial: matrix clauses facilitated trial times more for DP-pied-piping structures ($d = 906\text{ms}$) than for PP pied-piping structures ($d = 118\text{ms}$), $t = -2.72$, $p < 0.01$. When combined with

the acceptability judgments, the overall picture that emerges is that not only were DP pied-piping structures facilitated by the matrix clauses with respect to acceptability ratings, they were also facilitated with respect to trial time. Still, we hesitate to speculate about why the DP pied-piping structures were so greatly facilitated, as the pattern might be consistent with several different processing models.

While the crowd-sourcing methodology in psycholinguistics has previously shown identical results to results obtained from more traditional lab-based research (e.g., Munro et al., 2010; Sprouse, 2011; among others), it is still relatively new and little is known about what types of results generalize to other populations. Thus, despite all of our controls, we felt that we needed to replicate our core results with different subjects. Our second population was comprised of 88 undergraduate students at the University of Massachusetts Amherst, who completed the questionnaire online for course credit. We instituted the same procedure for excluding participants described above. Thirty participants were identified as non-native English speakers, duplicates, or simply inattentive, and removed from the analysis; another 15 were removed for counterbalancing purposes. The final dataset for replication consisted of 44 participants. Outliers were removed as before, with a comparable rate of data loss.

The second study indeed replicated the results from the first. Means and standard errors are provided in Figure 4 below. The statistical patterns in the acceptability rating data are in keeping with patterns obtained from the first group. In particular, participants rated pied piping in subordinate clauses ($M = 4.43$, $SE = 0.13$) significantly worse than pied-piping in matrix clauses ($M = 5.39$, $SE = 0.11$), $t = -3.93$, $p < 0.001$, consistent with the idea that the additional complexity of the subordinate clause introduces an independent cost. There was also a main effect of pied-piped element: participants rated structures with pied-piped PPs ($M = 3.80$, $SE = 0.11$) significantly worse than those with pied-piped DPs ($M = 6.02$, $SE = 0.09$), $t = -11.12$, $p < 0.001$. Lastly, an interaction between the two planned factors was observed, such that the effect of a subordinate clause on acceptability ratings was greater when the Pied-piped element was a PP ($d = 1.23$) than when it was a DP ($d = 0.70$), $t = -2.10$, $p < 0.05$. And, as before, this model fit the data better than a simpler model obtained from removing the interaction term, $\chi^2 = 4.41$, $p < 0.05$.

Interestingly, we did not replicate the trial time results from the AMT group: the only significant effect was that of clause type, in which subordinate clauses, as expected, elicited longer trial times than matrix counterparts. This is, as before, not clearly interesting, as items containing subordinate clauses were longer and more complex than items with only matrix clauses.



	DP-pied-piping	PP-pied-piping	Mean
Matrix	6.37 (0.10)	4.41 (0.15)	5.39 (0.11)
Subordinate	5.67 (0.15)	3.18 (0.15)	4.43 (0.13)
	6.02 (0.09)	3.80 (0.11)	

Figure 4: Mean acceptability judgment ratings (standard errors in parentheses) by condition for the UMASS group in Experiment 2.

3.4 Discussion

As noted above, the results of Experiment 2 support the prediction presented in (7b); the effect of clause type was significantly stronger in cases of PP-pied-piping than in cases of DP-pied-piping. This supports the notion that

PP-pied-piping is grammatically distinct from other pied-piping structures in English, in that it is essentially restricted to matrix environments, and therefore patterns with other cases of ‘massive pied-piping’.

However, another explanation of these effects is imaginable. As noted earlier, all cases of ‘non-massive’ pied-piping are also not optional. That is, in English there is a potential confound between pied-piping being ‘massive’ (in the technical sense) and it being optional. Thus, the special interaction between clause-type and pied-piping found for PPs might be due to its optionality rather than its ‘massiveness’. On the other hand, it is currently unclear why the optionality of a pied-piping structure would lead to the effects witnessed above, while both Heck (2008) and Cable (2010) provide formal syntactic explanations for why ‘massive pied-piping’ should be restricted to matrix environments.

3.5 Conclusions

The experiments described above sought to provide a more rigorous test of the predictions in (7), repeated below.

(15) Experimental Predictions to be Tested

- a. Pied-piping of PPs is less acceptable in English subordinate clauses than it is in main clauses.
- b. The contrast between pied-piping in main clauses vs. subordinate clauses is seen only for PP pied-piping. A comparable contrast is not found for pied-piping of DPs by *wh*-determiners.

As reported above, Experiment 1 supported the prediction in (15a), while Experiment 2 supported that in (15b). The general picture that emerges is that the effect of subordination on acceptability ratings of structures with pied-piped PPs is greater than the effect of subordination on structures with either preposition stranding (Experiment 1) or pied-piped DPs (Experiment 2). With this experimental support, we can be more confident in the raw judgment data reported by both Heck (2008) and Cable (2010), as well as in their conclusion that pied-piping of PPs in English occupies a special status amongst pied-piping constructions. Unlike pied-piping by *wh*-determiners or possessors, pied-piping of PPs is restricted to main clauses, and thus should be categorized as ‘massive pied-piping.’ We interpret the results as an initial corroboration of the idea that pied-piping of PPs is thus actually

‘illicit’ in English, its ill-formedness simply being mitigated in matrix environments as a result of independent grammatical factors.

References

- Allen, Cynthia (1980). *Topics in Diachronic English Syntax*. Garland Publishing, New York.
- Baayen, R. Harald (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics*. Cambridge University Press.
- Baayen, R. Harald, Douglas J. Davidson & Douglas M. Bates (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59:4, pp. 390–412.
- Bates, Douglas & Martin Maechler (2009). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-31.
- Cable, Seth (2010). *The Grammar of Q: Q-Particles, Wh-Movement and Pied-Piping*. Oxford University Press.
- Harris, Jesse A. & Christopher Potts (to appear). Predicting perspectival orientation for appositives. *Proceedings from the 45th Annual Meeting of the Chicago Linguistic Society*, Chicago Linguistic Society, Chicago, IL, vol. 45.
- Heck, Fabian (2008). *On Pied-Piping: Wh-Movement and Beyond*. Studies in Generative Grammar 98, Mouton de Gruyter.
- Munro, Robert, Steven Bethard, Victor Kuperman, Vicky Tzuyin Lai, Robin Melnick, Christopher Potts, Tyler Schnoebelen & Harry Tily (2010). Crowdsourcing and language studies: The new generation of linguistic data. *NAACL 2010 Workshop on Creating Speech and Language Data With Amazon’s Mechanical Turk*, ACL, Los Angeles.
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ross, John Robert (1967). *Constraints on variables in syntax*. Ph.D. thesis, MIT, Cambridge, MA.
- Sprouse, Jon (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods* :43.

Appendix A: Experiment 1 materials

1. a) Who did he dance with?
 b) I wonder who he danced with.
 c) With whom did he dance?
 d) I wonder with whom he danced.
2. a) Who did she talk to?
 b) I wonder who she talked to.
 c) To whom did she talk?
 d) I wonder to whom she talked.
3. a) Who did he work for?
 b) I wonder who he worked for.
 c) For whom did he work?
 d) I wonder for whom he worked.
4. a) Who did she smile at?
 b) I wonder who she smiled at.
 c) At whom did she smile?
 d) I wonder at whom she smiled.
5. a) Who did he travel with?
 b) I wonder who he traveled with.
 c) With whom did he travel?
 d) I wonder with whom he traveled.
6. a) Who did she sing to?
 b) I wonder to whom she sang.
 c) To whom did she sing?
 d) I wonder to whom she sang.
7. a) Who did he yell at?
 b) I wonder who he yelled at.
 c) At whom did he yell?
 d) I wonder at whom he yelled.
8. a) Who did she cook for?
 b) I wonder who she cooked for.
 c) For whom did she cook?
 d) I wonder for whom she cooked.

Appendix B: Experiment 2 materials

1. a) About what did she write?
 b) It's obvious about what she wrote.
 c) Which book did she write?
 d) It's obvious which book she wrote.
2. a) With what does Bill work?
 b) I asked with what Bill works.
 c) Which register does Bill work?
 d) I asked which register Bill works.
3. a) On what did Sarah cook?
 b) They wonder on what Sarah cooked.
 c) Which soup did Sarah cook?
 d) They wonder which soup Sarah cooked.
4. a) Near what did Lou perform?
 b) Kate knows near what Lou performed.
 c) Which act did Lou perform?
 d) Kate knows which act Lou performed.
5. a) From what did Bill run?
 b) I know from what Bill ran.
 c) Which race did Bill run?
 d) I know which race Bill ran.
6. a) With what did he draw?
 b) Fran asked with what he drew.
 c) Which picture did he draw?
 d) Fran asked which picture he drew.
7. a) With what did John doodle?
 b) I wonder with what John doodled.
 c) Which picture did John doodle?
 d) I wonder which picture John doodled.

8. a) About what did Joan mumble?
 b) It's obvious about what Joan mumbled.
 c) Which excuse did Joan mumble?
 d) It's obvious which excuse Joan mumbled.
9. a) From what did Peter drink?
 b) Barb wonders from what Peter drank.
 c) Which wine did Peter drink?
 d) Barb wonders which wine Peter drank.
10. a) From where did she escape?
 b) I know from where she escaped.
 c) Which prison did she escape?
 d) I know which prison she escaped.
11. a) With what did she eat?
 b) He wonders with what she ate.
 c) Which dish did she eat?
 d) He wonders which dish she ate.
12. a) Towards what did Bill drive?
 b) I asked towards what Bill drove.
 c) Which car did Bill drive?
 d) I asked which car Bill drove.

Catch items. Participants saw each of the following four items, randomly dispersed within the experiment. Participants who rated any one of these items a 3 or above on the 7-point scale were excluded from analysis.

1. Most book did Lance enjoy in.
2. More scooters than bicycles students has ridden.
3. Doris talked a book a page.
4. I think he remembers he.

Dynamic interpretation: Finding an antecedent for VPE

Lyn Frazier and Charles Clifton, Jr.

University of Massachusetts Amherst

Abstract

When the potential antecedents for an elided verb phrase (VP) appear in a different sentence than the elided VP, comprehenders tend to choose an antecedent that is part of the main assertion of the antecedent sentence (Frazier & Clifton 2005) presumably because of its salience in the discourse representation (Kamp & Reyle 1993). Here we report a study exploring this effect further, showing that a conjoined VP, rather than a single VP conjunct, was the preferred antecedent when an antecedent sentence contained either conjoined VPs or conjoined clauses, but less often when the same antecedent content was expressed in a series of sentences. Remarkably, ‘conjoined verb phrase’ interpretations were still given roughly half the time when verb phrases occurred in separate sentences if those sentences were parallel in terms of tense and aspect, and related to each other in terms of content. We view these results as an initial attempt to explore ‘dynamic interpretation,’ i.e., how processes that affect the construction and use of discourse representations can influence the interpretation of sentences.

1. Introduction

Speakers and authors do not place information randomly in the sentences they produce. The linear and structural placement of constituents is influenced by information structure properties like topic, focus, discourse-old and discourse-new status (Birner & Ward 1998; Prince 1981; Vallduvi 1992) and, we have proposed, the notion of main assertion (Frazier & Clifton 2005). It is not random whether clause A is couched as a subordinate/relative/embedded clause and clause B as the main clause or vice versa. This depends on what the speaker intends and what the discourse is about.

Information structure and syntactic structure are both important for ellipsis. The antecedent for an elided constituent must be given (e.g., Merchant, 2001). In prior work (Frazier & Clifton 2005) we have shown that, when the antecedent for an elided verb phrase (VP) can be found inside the current sentence, the preferred antecedent is often the antecedent lower in the syntactic tree. For instance, in (1a), the preferred antecedent of *Mary did too* is *went to Europe*. However, when the antecedent must be found outside the current sentence, the preferred antecedent shifts and is more likely to be part of the main assertion of the preceding sentence (Frazier & Clifton, 2005 formulated this claim as the Main Assertion hypothesis). For example, readers and listeners chose the matrix antecedent (*said Fred went to Europe*) more often in (1b) than the did in (1a). They also chose the VP of the main clause as an antecedent more often than the VP of the subordinate clause in (1) regardless of clause order.

- (1) a. John said Fred went to Europe and Mary did too.
b. John said Fred went to Europe. Mary did too.
- (2) a. Mary laughed after she made a joke about the supervisor. Then Tina did too.
b. After Mary laughed she made a joke about the supervisor. Then Tina did too.

The idea that information structure is what's at issue with the main assertion preference is reinforced by the finding that fewer matrix clause antecedents are chosen when the matrix clause could be interpreted as an epistemic clause, which only expresses the speaker's degree of commitment to the assertion as in (3a), than in clauses with third person subjects which are not open to an epistemic interpretation (3b). In a control study with verbs like *announce*, which are not open to an epistemic interpretation, there was no difference between first person (3a) and third person (3b) subjects.

- (3) a. I think Mary smokes. Sam does too.
b. The teacher thinks Mary smokes. Sam does too.

The main assertion made by a speaker by using an utterance typically involves the content of constituents that occur high in the syntactic tree (e.g., the main clause, not a relative clause). The main assertion must be conveyed

by material which is not presupposed. Further, the content of the main assertion is entailed. Typically, the main assertion will contain material that is focused.

Much of our previous work on ellipsis has stressed the necessity of finding a syntactically appropriate antecedent for the elided phrase. Frazier & Clifton (2001) argued that the syntactic form (more precisely, the logical form, LF) of an antecedent is copied into the ellipsis to interpret the latter. Arregui et al. (2006) showed that listeners and readers can often cope with elided phrases that do not have syntactically perfect ('matching,' certain morphological details aside) antecedents, as in the attested example in 4, taken from Hardt (1993). However, they suggested that VP ellipsis (VPE) sentences without matching antecedents are acceptable only if the listener/reader has sufficient evidence to patch up the antecedent at LF, converting it into an appropriate syntactic form.¹ Frazier (2006) showed that this preference for a syntactically-matching antecedent held across different discourse relations (e.g., causality vs. temporal contiguity) that have been claimed (Kehler 2000, 2002) to affect the acceptability of elided phrases.

- (4) This information could have been released by Gorbachov, but he chose not to [release this information]

Our previous work thus shows that the form, not simply the truth-conditional properties, of a phrase constrains its acceptability as an antecedent of an ellipsis. However, some of our previous work has suggested

¹ Arregui et al. (2006) suggested that a VPE without a matching antecedent may be more acceptable if the example sounds like a likely speech error (syntactic blend). In syntactic blends, a speaker is more likely to substitute an active elliptical phrase for an intended passive than vice versa; in acceptability judgments, a reader/listener finds an active ellipsis with a passive antecedent more acceptable than the other way around. Similarly, with syntactic blends we have found in a small study with two items like (i) that the a-form is rated more acceptable than the b-form. In the a-form one can readily imagine the motivation for changing impersonal you to impersonal we for politeness reasons, in order to avoid any implication that might arise in (ib) that the addressee is delusional.

- (i) a. If you think this is going to solve the terrible problem in Najaf, we're deluding ourselves.
b. If we think this is going to solve the terrible problem in Najaf, you're deluding yourself.

(48 subjects saw one a-form and one b-form of two sentences like (i) in a written questionnaire, 1=Unacceptable to 5=Acceptable; $a = 4.05$, $b = 3.37$, $t(46) = 2.73$, $p < .01$) The idea that the likelihood of a speaker making a particular type of syntactic blend influences the acceptability of the sentence fits with our approach to flawed antecedents as being sometimes acceptable though still ungrammatical.

that this formal constraint can be relaxed when an antecedent is found in the discourse representation. For instance, Experiments 8a and 8b in Frazier & Clifton (2005) suggested that sluicing sentences like *They hired someone who won something but I don't know what* could be interpreted by the discourse processor even though the antecedent of the sluice was contained within a relative clause island, violating a formal constraint.

In the present paper, we begin to generalize this claim, proposing what we call “dynamic interpretation” of ellipsis. The claim is simply that, while choice of an antecedent of an ellipsis is strongly affected by the form of the antecedent when the form is highly available in memory (as it presumably is within the current sentence), the appropriateness of an antecedent can change dynamically as a discourse progresses. This change can be as simple as the precise memory for the form of an antecedent fading as a syntactic boundary is crossed, or as complicated as a change in the information structure properties (e.g., given vs new) of the antecedent changing as the form of the discourse changes. The following experiment represents our initial exploration into phenomena of dynamic interpretation of ellipsis.

2. Experiment

We presented short discourses, ending with a VPE sentence, in a self-paced reading task. Its primary goal was to determine how the VPE is interpreted in different discourse contexts. To this end, each discourse was followed by a two-choice question about the VPE. The primary data of the experiment consisted of the answers to these questions.

The discourses are illustrated in (5). Consider (5a), with a conjoined VP (*travels in winter and stays home in summer*) in the first sentence.

- (5) a. Ian travels in winter and stays home in summer. George does too.
- b. Ian travels in winter and he stays home in summer. George does too.
- c. Ian travels in winter. He stays home in summer. George does too.
- d. Ian used to travel in winter. He stays home in summer. George does too.

We would expect that readers would be likely to report the larger antecedent (*George travels in winter and stays home in summer*) in as the antecedent of

the VPE sentence *George does too* both because of the parallelism between the two VPs (suggesting that the two should be treated alike) and because, following the Main Assertion hypothesis, discourse prominence will favor this antecedent. By contrast, in (5c), where the conjoined constituent has been separated into distinct sentences, parallelism of the two VPs would still favor the two-phrase antecedent. However, the sentence boundary may lead to salience differences between the first and second VP, assuming that the memory representation decays over time but and as a sentence boundary is encountered (so that constituents in the sentence most recently processed will be more accessible). At the end of the second clause, the first clause will have decayed more when it is in a separate sentence than when it is the first clause of conjoined sentences. Consequently, for salience or memory accessibility reasons, just the recent VP will be favored as antecedent in (5c). In (5d), due to the difference (nonparallelism) in the syntax and in the ‘aspect’ of the first two sentences, we expect an even stronger preference for the recent VP to be selected as antecedent. In short, we expect more large antecedents in (5a) > (5c) > (5d).

The predictions concerning (5b) depend entirely on what one assumes about the syntax of these sentences. For the moment, let’s assume that there are two full clauses conjoined in (5b). Then we might expect fewer conjoined VP interpretations in (5b) than in (5a) because in (5b) deriving a conjoined VP would involve an extra operation not needed in (5a) where a conjoined VP antecedent already exists. Alternatives will be considered in the discussion.

3. Methods

3.1 Materials

To test the above predictions, we constructed sixteen examples like (5), with four versions of each. In the a-version, the first sentence superficially contained conjoined VPs and the second sentence contained an elided verb phrase. The b-form of the sentence was identical to the a-form except that the second clause of the first sentence contained a pronoun with features appropriate for the subject of the first clause. The c-form was identical to the b-form except that the two clauses of the first sentence in the b-form were presented as separate sentences in the c-form. In the d-form, the first sentence of the c-form was altered so that it contained “used to.” All sentences

are presented in the Appendix.

Each sentence was followed by a two-choice question of the form *What about George?* followed by two alternative interpretations: *travels in winter/used to travel in winter and stays home in summer*; *stays home in summer*. The first interpretation will be referred to as a two-phrase or long answer; the latter, as a one-phrase or short answer. The one-phrase answer always presented the more recent of the two possible VP antecedents, under the assumption that choice of the more distant antecedent is unlikely and arguably ungrammatical.

These 16 sentences were combined with a total of 88 filler discourses and discourses from other experiments, half of which were followed by two-choice questions. A practice list, with 8 short discourses (4 followed by questions) was also constructed.

3.2 Participants and procedures.

Forty-eight University of Massachusetts undergraduates were tested individually in half-hour sessions, for course credit. After receiving instructions, each participant was seated in a sound-deadened chamber containing a video terminal and tested on the practice list. The participant first saw an array of underscore marks on the video terminal, one for each letter of the discourse. As soon as the participant pulled a lever with the right hand, the first sentence (or for conditions (5c) and (5d), the first two sentences) appeared on the screen, replacing the underscores. The participant read the material on the screen and pulled the lever, making the material disappear and the final VPE sentence appear. After the participant read this sentence and pulled the trigger, the discourse disappeared and was replaced by a question positioned above the two possible answers. Half the time the short, one-phrase, answer was on the left and the longer, two-phrase answer on the right; the other half of the time, position was reversed. The participant was to pull a lever under the answer that s/he intuitively thought that the VPE sentence meant. Reading time for each presentation segment and question-answering time and choice were recorded. However, since the experiment was designed to determine the interpretation of ellipses in different discourses, only the actual answer choices will be analyzed. (Note that the presentation procedure differed for the other sentences being tested, some of which were presented all at once, with others being presented in

segments consisting of two or three words.)

The four forms of each of the 16 experimental discourses were assigned to four different presentation lists using a latin square counterbalancing procedure that ensured that each participant read four discourses in each form, and that across all participants, each sentence was tested equally often in each form. Equal numbers of participants saw each of the four forms, with the discourses presented in individually-randomized orders.

3.3 Results

Table 1 presents the percentages with which the two-phrase *travels in winter and stays home in summer* answer was chosen (and the complementary percentage of choice of the simple, *stays home in summer* answer).

	Condition	Percentage Choice		Question-Answer Time (ms)
		Two-phrase	One-phrase	
5a.	Conjoined VP	84	16	1112
5b.	Conjoined clause	86	14	1024
5c.	Two clause, parallel	55	45	1187
5d.	Two clause, non-parallel	32	68	1135

Table 1: Percentages of Answer Choices and Question-answering Times for Experiment 1.

The difference among the four choice percentages was highly significant ($F_1(3, 141) = 54.73, p < .001$; $F_2(3, 45) = 83.53, p < .001$). Each pairwise difference similarly highly significant, except for the difference between (5a) and (5b).

3.4 Discussion

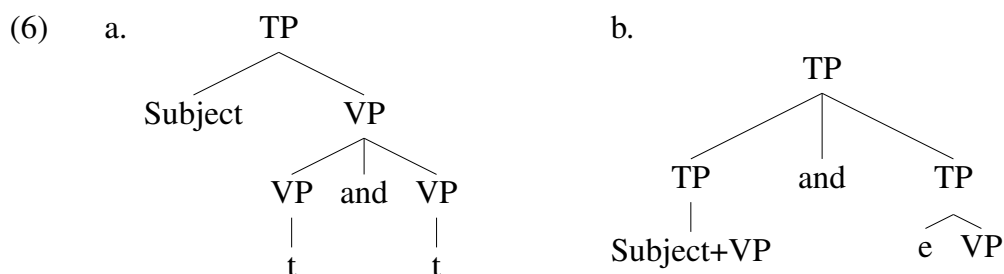
As predicted, the conjoined VP interpretation was preferred over the recent VP antecedent interpretation in (5a). This is expected given that conjoining the two VPs in (5) presumably gives them equal discourse prominence. Interestingly, (5b) received as many conjoined VP interpretations as (5a). This was not expected, at least not given a syntax where the structure of (5b) involves coordination of full clauses. In (5c) readers preferred the VP of the more recent sentence as the antecedent slightly less than half the time.

However, in (5d) the recent VP antecedent was strongly preferred once the parallelism in the tense/aspect of the two VPs was eliminated.

Setting aside (5b) temporarily, these results are interesting for a variety of reasons. The preference for a conjoined VP interpretation, while not terribly surprising, does provide further evidence that across sentence boundaries information-structure notions, not locality, determine the salience of constituents. (5c) shows that ‘locality/recency’ can still play a role in the discourse representation, but it appears that sentence boundaries influence where preferred antecedents are found, presumably because they influence the accessibility of material in memory. Probably on any theory, the fact that fewer conjoined VP antecedents were chosen when the VPs were in separate sentences and had distinct tense/aspects is expected: memory accessibility factors and parallelism factors will mitigate against the conjoined interpretation.

Turning to (5b), the question is why it received just as many conjoined VP interpretations as (5a). One possibility in principle is that the sentence form doesn't matter for intersentential ellipsis. Logically equivalent antecedents might be processed in just the same way, independent of syntactic form. But we have already reported that this approach will not suffice: the relation between a passive and active is the same as the relation between an active and a passive, yet their acceptability differs (Arregui et al. 2006). So truth-conditional or logical equivalence will not suffice to characterize acceptable antecedents.

But imagine that the surface syntax of conjoined VPs and same-subject conjoined clauses are the same, possibly conjoined VPs with a VP-internal subject trace, as shown in (6a)(see Burton & Grimshaw, 1992 and McNally, 1992), or conjoined at a higher level such as TP (or vP), as illustrated in (6b):



If we accept the analysis in (6a), then in (5a), the “conjoined VP” sentence, we would claim that the traces left in the VP-internal subject position are silent. In contrast, in (5b), the “conjoined clause” sentence, the trace in the second VP would be spelled-out, e.g., treated as a resumptive pronoun. If this is the correct structure of (5b), then of course, (5b) should pattern with (5a) receiving mostly conjoined VP interpretations.

Additional evidence in favor of the structure in (6a) comes from the behavior of *both*, which typically is acceptable with conjoined VPs but not with conjoined clauses, as illustrated in 7.

- (7) a. Fred both loves Mary and hates Trish.
 b. Fred both loves Mary and he hates Trish.
 c. ?? Fred both loves Mary and the bastard hates Trish.

This syntactic account of the behavior of (5b) predicts that readers should not favor a conjoined VP analysis in (8a), where the subject of the second conjunct must be outside the VP in order to precede the modal. Comparing (8a) to (8b), where the subject could be inside the VP, intuitions do suggest fewer conjoined VP responses in (8a).

- (8) a. Josh will go to Paris and he could visit Madrid. Fred would too.
 b. Josh has gone to Paris and he visited Madrid. Fred will too.

Another possible analysis is one with a TP (a ‘tense phrase’) in both conjuncts, along the lines that was illustrated in (6b). Given that the verb in the second conjunct is tensed, it is not unlikely that the verb is in a TP which could host tense (see Van Valin, 1986). This suggests that some empty category is serving as subject of the second conjunct even when the conjunct does not contain an overt pronominal subject. (See Rögnvaldsson, 1993, for evidence from Icelandic suggesting that something along the lines of (6b) is required for a general solution.) Both sorts of analyses would attribute the ‘conjoined VP’ and conjoined clause sentences comparable structures and thus explain why (5a) and (5b) were treated identically in terms of their preferred interpretations. Of course there are far-reaching syntactic consequences to this proposal which we will treat as being outside the scope of this paper (though clearly they are relevant to assessing the adequacy of this explanation for the results concerning (5b)).

Turning to (5c) and the surprisingly large number of ‘conjoined’ antecedent responses obtained even when the verb phrases were presented in separate sentences, one might assume that Predicate conjunction is available at LF and it may apply freely when the subjects of two predicates refer to the same discourse referent. Presumably Predicate conjunction would be most likely to apply when the predicates are related as part of a single discourse unit and thus primarily only when the predicates share the same tense and aspect, thereby accounting for why (5c) received more ‘conjoined’ responses than (5d).

In terms of operations performed on antecedents at LF, we have argued elsewhere (Arregui et al. 2006; Frazier 2008) that such operations, or “recycling,” are most likely when the actual form of the input and the desired form of the antecedent for an ellipsis are systematic alternative ways of expressing a message. The idea is that speakers make syntactic blends when they produce language. A speaker may have formulated an antecedent clause as a passive, but then misremember and formulate an ellipsis clause as an active. Listeners may find such forms relatively acceptable when they are easy to repair and they correspond to forms speakers are likely to produce. Similarly, a speaker may convey a series of events using conjoined verb phrases, conjoined clauses or separate sentences and then forget which form was actually used when an ellipsis clause is uttered. If the listener can reconstruct the speaker’s intent and can make the appropriate predicate conjunction at LF, then an elided verb phrase might be interpreted with an antecedent that was not actually expressed as a verb phrase (*though it could have been!*).

Of course verb phrase conjunction, clausal conjunction and their separate sentence counterparts (e.g., (5a,b,c)) do not always convey precisely the same information, e.g., support the same implicatures, or invite the same assumptions about the number and chronological order of events. But in ellipsis, whatever interpretation has been assigned to the antecedent seems to carry over to the elided clause; it is extremely difficult if not impossible to force non-parallel interpretations of the antecedent clause and the ellipsis clause, e.g., to interpret *John has three children and so does Bill* as John has exactly three children and Bill has three or more children. Presumably this parallel interpretation of antecedent clause and ellipsis clause means that potential interpretive differences between conjoined verb phrases, conjoined

clauses, and separate sentences would not be a reason to block the listener from accepting a ‘conjoined’ antecedent interpretation of an elided verb phrase even in the absence of a conjoined verb phrase in the surface syntax.

4. Conclusions

Interpreting language as it arrives involves constant reference to a context that is changing as new input is encountered. Here we have investigated one instance of how the current form of the syntactic and discourse representation changes as a function of the syntax of the input, namely, whether it is presented as one sentence or two. In ongoing work, we are studying effects of the semantic type of the input, whether it is part of the at-issue content or a supplement, and the discourse updating, involving inferences allowing an antecedent clause to drop out if asserted information entails its truth. With respect to sentence boundaries, the present work and earlier work suggests that the syntax of a sentence will decay when the reader or listener moves on to a new sentence. Presumably the syntax of the old sentence is allowed to decay when the syntax of that sentence is no longer relevant for current processing, e.g., syntactic structuring of new input. By contrast, the reader or listener has a more accessible syntactic representation of a clause when it is part of the current sentence, where its structure is more likely to be needed to structure further input.

The issue of dynamically updating context is a central one in contemporary theories of semantics. Using psycholinguistic techniques to look at ongoing changes in the representations used for various processing tasks seems like a complementary and necessary part of an explicit theory of language interpretation. Theories like Discourse Representation Theory may offer a general picture of the representations available to language users. Identifying empirical constraints on how and when those representations are updated will presumably allow a more detailed picture of how and why listeners/readers comprehend discourse the way they do.

Acknowledgment. We are very grateful to Rajesh Bhatt, Kyle Johnson, Alan Munn, and Chris Potts for helpful discussion of this research. The work was supported by NIH Grant HD-18708 and NSF Grant BCS-0090694 to the University of Massachusetts.

References

- Arregui, A., C. J. Clifton, L. Frazier & K Moulton (2006). Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language* :55, pp. 232–246.
- Birner, B. & G. Ward (1998). *Information status and noncanonical word order in English*. John Benjamins, Amsterdam.
- Burton, S. & J. Grimshaw (1992). Coordination and VP-Internal subjects. *Linguistic Inquiry* :23, pp. 305–313.
- Frazier, & Clifton C., L. (2006). Ellipsis and discourse coherence. *Linguistics and Philosophy* 29, pp. 315–346.
- Frazier, L. (2008). Processing ellipsis: A processing solution to the under-generation problem. Chang, Charles & Hannah Haynie (eds.), *Proceedings from the West Coast Conference on Formal Linguistics*, Cascadia Press, vol. 26.
- Frazier, L. & C. Clifton (2001). Parsing coordinates and ellipsis. Copy α . *Syntax* 4, pp. 1–22.
- Frazier, L. & C. Clifton (2005). The syntax-discourse divide: Processing ellipsis. *Syntax* 8, pp. 121–174.
- Hardt, D. (1993). *Verb phrase ellipsis: Form, meaning and processing*. Ph.D. thesis, University of Pennsylvania.
- Kamp, H. & U. Reyle (1993). *From discourse to logic: Introduction to model theoretic semantics of natural language, formal language and discourse representation theory*. Kluwer Academic Publishers, Dordrecht.
- Kehler, A. (2000). Coherence and the resolution of ellipsis. *Linguistics and Philosophy* 23, pp. 533–575.
- Kehler, A. (2002). *Coherence, Reference and the Theory of Grammar*. CSLI Publications, Stanford.
- McNally, L. (1992). VP-coordination and the VP-Internal subject hypothesis. *Linguistic Inquiry* 23, pp. 336–341.
- Merchant, J. (2001). *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. Oxford University Press, Oxford.
- Prince, E. (1981). Towards a taxonomy of given/new information. Cole, Peter (ed.), *Radical pragmatics*, Academic Press, New York, pp. 223–254.

- Rögnvaldsson, E. (1993). Coordination, atb-extractions, and the identification of *pro*. *Harvard Working papers*, Harvard, vol. 3.
- Vallduvi, E (1992). *The information component*. Garland, New York.
- Van Valin, Jr., Robert D. (1986). An empty category as the subject of tensed S in English. *Linguistics Inquiry* 17, pp. 581–586.

Appendix

Materials used in Experiment 1. Full paradigm illustrated for Item 1.

1. a) I like Mary and I hate Sam. Fred does too.
 b) I like Mary and hate Sam. Fred does too.
 c) I like Mary. I hate Sam. Fred does too.
 d) I used to like Mary. I hate Sam. Fred does too.
2. Ian travels in winter and he stays home in summer. George does too.
3. Jessica eats cheese and she eats fish. Annie does too.
4. Gloria waters the plants and she feeds the cats. Stacie does too.
5. Will studies Psychology and he studies Neurology. Ren does too.
6. Jake plays chess and he plays mahjong. Rick does too.
7. Brenda bakes this weekend and she cleans house. Lydia does too.
8. Nathaniel writes poetry and he writes short stories. Lisa does too.
9. Patrick sings arias and he plays ballads. Glen does too.
10. Ernie ice skates and he golfs. Bill does too.
11. Lucinda babysits and she sells newspapers. Gail does too.
12. Barbara goes bird-watching and she collects butterflies. Ellen does too.
13. Angela flies small planes and she parachutes. Elizabeth does too.
14. Emily knits and she sews. Paula does too.
15. Pamela studies mathematics and she repairs cars. Bella does too.
16. Tim annoys Australians and he amuses Americans. Ken does too.

D-linking and Memory Retrieval: The Annoying Case of Sluicing

Lyn Frazier and Charles Clifton, Jr.

University of Massachusetts Amherst

1. Introduction

In contrast to bare interrogative pronouns like *who*, interrogative phrases like *which guy* have been analyzed as presupposing the existence in the discourse context of a set of entities of the type denoted by the nominal (e.g., a set of guys) (Pesetsky 1987; Enç 1991). Such phrases have been termed discourse-linked (d-linked). Linguists have argued that d-linking affects the acceptability of sentences. They have observed that interrogative *wh*-phrases cannot generally be extracted from complex noun phrases, as illustrated in (1a). However, they have also observed that the unacceptability of such extraction is reduced if the interrogative phrase is discourse-linked (d-linked), as in (1b).

- (1) a. *Who did you meet the girl who dated?
b. (*)Which guy did you meet the girl who dated?

The contrast between d-linked and non-d-linked interrogatives has also been shown to have processing consequences. For instance, Frazier & Clifton (2002) showed that a d-linked phrase was preferred as an antecedent of a pronoun, compared to a non-d-linked phrase, arguing that a d-linked phrase sets up a discourse entity, the preferred antecedent of a pronoun (Garnham et al. 1995) (see below for further discussion and additional demonstrations).

Recently Hofmeister (2009a) put forward an interesting and plausible hypothesis that might provide an account of observed differences between d-linked and non-d-linked interrogatives. Rather than attributing these differences to differences in the existence of discourse entities, he proposed

that more complex interrogative phrases are simply easier to retrieve from memory than their less complex counterparts. Specifically, he proposed the Representational complexity hypothesis in (2).

- (2) Representational complexity hypothesis (Hofmeister 2009a)
- A. Given two linguistic phrases that can each felicitously describe some discourse entity *e*, the phrase that is syntactically and/or semantically more complex will facilitate the retrieval process initiated at the next (overt or covert) reference to *e*.
 - B. An expression x_1 is more complex than an expression x_2 if the semantic and syntactic information encoded by x_2 is a proper subset of the information encoded by x_1 .

One might worry about the precise formulation of the complexity hypothesis. It seems not to apply to many intuitively relevant examples (*man* and *graduate student* are not in a subset relation). It also may not apply to critically relevant examples (*who* vs. *which* or *which person*) depending on the precise feature complex assumed for interrogatives. Nonetheless, we accept that the hypothesis claims that d-linked wh-phrases like *which guy* are more complex than non-d-linked phrases like *who*, and that it could account for the effects described above.

The concern of the present paper is whether memory retrieval is in fact necessarily involved in instances where a d-linked wh-phrase facilitates comprehension. This question is a general one involving wh-dependencies of all sorts, whether they involve islands or not. We report an experiment comparing complex (d-linked) and simple (indefinite) interrogatives in sluicing sentences. The results draw into question the basic idea that the advantage for d-linked phrases should be attributed solely to memory retrieval.

The basic hypothesis that Hofmeister is proposing (in (2) and in very slightly different forms in related work) is plausible. The idea is that the complexity of encoding a phrase influences how easy it is to later retrieve that phrase from memory. This fits with a long tradition in memory research known as 'depth of processing' effects: the more deeply material is processed at encoding, the easier it is to recall later on (Craik & Lockhart 1972: *inter alia*). In Hofmeister's own conception, the idea is that "The process of building the complex representation creates a highly activated

memory network and the ensuing cognitive structure allows for multiple retrieval strategies. Put slightly differently, it is both the process and result of building a complex representation that facilitates the retrieval process.” (Hofmeister 2009a: page 11)

The evidence Hofmeister presents in support of (2) derives from a series of acceptability judgment and self-paced reading studies. Hofmeister (2009a) (Experiment 1), tested sentences like (3).

- (3) a. simple: It was a communist who the members of the club banned from ever entering the premises.
- b. mid: It was an alleged communist who the members of the club banned from ever entering the premises.
- c. complex: It was an alleged Venezuelan communist who the members of the club banned from ever entering the premises.

At the clefted constituent, (log residual) reading times were longer for the complex and mid phrases than for the simple phrase; at the first word after the gap following *banned*, the complex phrase condition was read faster than the mid and the simple conditions. Ignoring the intermediate status of the mid condition, the results support the Representational complexity hypothesis: retrieval of the more complex phrase appears to be easier at the gap. Experiment 2 in Hofmeister (2009a) made a similar point, although his Experiment 3 indicated that at least some of the apparent effect of complexity could be due to plausibility or co-occurrence frequency.

In similar studies investigating the self-paced reading of questions with d-linked and indefinite interrogatives, Hofmeister (2009b) found evidence for faster reading following an extraction site when the wh-phrase was d-linked than when it was not. One experiment examined reading time of questions at and after an extraction site that was inside an island (*dismissed* in (4)), and found slower reading when the extracted element was a bare interrogative (4a) than when it was a d-linked wh-phrase (4b), which did not differ from the fully grammatical baseline condition (4c). A second experiment found a similar result for sentences without islands: Reading at and after the gap was slower for a bare interrogative wh-phrase (5a) than for a d-linked wh-phrase (5b), which did not differ from a more complex d-linked wh-phrase (5c).

- (4) Leadin Sentence: Albert learned that the managers dismissed the employee with poor sales after the annual performance review.
 - a. bare: Who did Albert learn whether they dismissed (*e) after the annual performance review?
 - b. which: Which employee did Albert learn whether they dismissed (*e) after the annual performance review?
 - c. baseline: Who did Albert learn that they dismissed (e) after the annual performance review?
- (5) Scooter hid from the reporter who talked about the recent ABC political poll on a recent evening news segment.
 - a. simple: What did the reporter that Scooter avoided discuss (e) during an evening news segment?
 - b. which: Which poll did the reporter that Scooter avoided discuss (e) during an evening news segment?
 - c. complex: Which political poll did the reporter that Scooter avoided discuss (e) during an evening news segment?

Sag et al. (2009) report similar selfpaced reading effects for bare and which-phrases extracted from complex noun phrases (e.g., *I saw who/which criminal Emma doubted the report that we had captured (*e) in the nationwide manhunt*). Reading time at the gap site was slower for bare interrogatives than for d-linked wh-phrases (an effect which actually appeared earlier in the sentence, at the complementizer). Acceptability rating studies showed that the d-linked which-sentences tested in self-paced reading are rated as more acceptable than their bare-wh-phrase counterparts.

The approach to d-linking that Hofmeister and colleagues (Hofmeister 2009a,b; Sag et al. 2009) have pursued bears on basic operations underlying human sentence comprehension, specifically, the retrieval of earlier constituents from memory. In terms of the larger picture, it is also interesting with respect to important issues in linguistic theory, such as whether locality conditions (Complex NP Constraint, subjacency) and Superiority are encoded in the grammar. Their proposal is that island violations such as that found in (1a) are simply the result of processing complexity: the complexity of retrieving simple phrases (*who*) from memory along with the complexity of processing long dependencies (though, to our knowledge,

they never address the vexing issue of whether two sources of complexity in language processing are additive or interactive – an issue that is currently unresolved).¹

For present purposes, our concern is whether it is true that filler-gap dependencies with d-linked fillers are easier to process than their non-d-linked counterparts only because of the ease of retrieving d-linked phrases from memory. To address this question we investigated the role of d-linking in sluicing, where no overt items intervene between the interrogative and its gap. On current views of memory, the last item to be processed remains active in memory and thus retrieval from memory should not be involved when a d-linked phrase must bind a gap without any intervening material (see review and integration of memory research in Jonides et al. (2008), for example).

2. Experiment

We report a written acceptability judgment experiment, in which participants rated the acceptability of sluicing sentences (see Chung et al., 1995; Merchant, 2001; Frazier & Clifton, 1998, for discussion of the structure and processing of sluicing). We compared the acceptability of indefinite/bare interrogative forms and d-linked forms, as illustrated in (6). We assume that the antecedent for the elided constituent is copied into the ellipsis site following *what/which one*, and a variable replaces *a new vehicle* in the elided clause. According to the Representational Complexity hypothesis, there is no reason to expect the d-linked (6b) form to have an advantage: no overt intervening material must be processed between the wh-phrase and the ellipsis site, so there should be no need to retrieve the wh-phrase from memory. However, if there is some other source of advantage for d-linked interrogatives (e.g., the existence of a discourse referent; see Frazier & Clifton, 2002 and discussion below), then we would expect the d-linking advantage to be present even when memory retrieval is not at issue.

¹ Presumably the prohibition against extracting from an island could be stated in the grammar and apply to both simple and d-linked/complex wh-phrases, it could be stated in the grammar and apply only to simple wh-phrases, or there could be no grammatical prohibition, only a processing constraint. While data like those presented in the Hofmeister (2009b) paper, and the present paper, may bear on this issue, further arguments are needed to choose between grammatical and processing accounts.

- (6) a. Britney likes this guy who destroyed a new vehicle but she didn't reveal what.
 b. Britney likes this guy who destroyed a new vehicle but she didn't reveal which vehicle.
 c. I know Britney destroyed a new vehicle but she didn't reveal what.
 d. I know Britney destroyed a new vehicle but she didn't reveal which vehicle.

2.1 Method

2.2 Materials

Sixteen sets of four sentences were constructed. Two of the four were sluice sentences that required extraction of the antecedent of the sluice from a relative clause ((6a) and (6b)). In the other two, the antecedent could be extracted from an accessible complement clause position ((6c) and (6d)). Crossed with this factor was the type of wh-phrase: indefinite bare interrogative ((6a) and (6c)) vs. d-linked ((6b) vs (6d)). The restrictor noun was contentful (*which vehicle*) in 12 of the 16 sentences and not contentful (*which one*) in four. All experimental items appear in the Appendix.

The 16 sentences were divided into four counterbalanced sets, with a different version of a sentence in each set, so that each set contained four sentences of each of the forms illustrated in (6). These 16 sentences were combined with 74 other sentences of a variety of forms. Each of the experimental sentences and all but 16 of the other sentences were followed by a 5-point rating scale, ranging from 1 (unacceptable) to 5 (acceptable). The remaining 16 sentences were followed by 2-choice interpretation questions. The 90 sentences were randomized using 12 different random orders and printed on paper forms that subjects were to complete by hand.

2.3 Participants and procedures

Forty-eight University of Massachusetts undergraduates were tested in individual sessions lasting under half an hour. They were given written instructions about the questionnaire, indicating that they were to rate a sentence as "acceptable" "as long as it is one you could easily imagine saying yourself or hearing and not noticing anything odd about it." They were told to indicate that the sentence is unacceptable "if it is one you would not say

except by mistake or one that you would not expect to hear a native speaker of English saying.”

2.4 Results

The mean ratings for the four sentence forms appear in Table 1. D-linked sentences were rated as more acceptable than non-d-linked sentences ($F_1(1, 47) = 24.02, p < .001; F_2(1, 15) = 15.04, p < .001$), and complement clause extraction site sentences were rated as more acceptable than relative clause extraction site sentences ($F_1(1, 47) = 26.63, p < .001; F_2(1, 15) = 33.06, p < .001$). The interaction between the two factors was not significant ($F_1(1, 47) = 1.35, p = .25; F_2(1, 15) = 1.33, p = .27$).

Wh-phrase	Extraction Site	
	Relative Clause	Complement Clause
non-d-linked	3.17	3.79
d-linked	3.69	4.13

Table 1: Mean Ratings (1= unacceptable, 5 = acceptable)

2.5 Discussion

The results suggest that, even in sluicing, examples with d-linked wh-phrases are rated more acceptable than their simple counterparts. In the sluicing examples tested, there is no need to retrieve a wh-phrase across overt intervening material. However, it is still possible that memory retrieval of some form is involved. We assume that the antecedent for the elided interrogative clause must be copied, or the structure of the antecedent must be shared (Frazier & Clifton 2005; Martin & McElree 2008). In (6c,d), for example, the antecedent is “Britney destroyed a new vehicle.” The indefinite must be replaced with a variable producing “Britney destroyed x.” Perhaps activating the antecedent clause is easier with a d-linked interrogative, as in (6b). The d-linked interrogative phrase may simply serve as a better retrieval cue.

Although interesting, this is not the Hofmeister proposal. His proposal concerns the retrieval of the wh-phrase itself. Further, the retrieval cue hypothesis put forward above does not really apply to the examples which lacked a contentful restrictor, i.e., the examples with “which one.”

Nevertheless, the examples without a contentful restrictor showed the same pattern as those with a contentful restrictor: which one: 3.75 for non-d-linked complement, 4.04 for d-linked complement, 3.14 for non-d-linked relative, 3.45 for d-linked relative versus 3.8, 4.15, 3.17, and 3.76 respectively, for the contentful restrictor examples. This suggests that neither the original retrieval hypothesis nor the ‘cue for retrieval of the antecedent’ hypothesis is correct, or at least, is not the full story. We explore an alternative below after taking up the other main effect in the experimental results.

The results of the experiment indicated that complement clause questions were rated more acceptable than relative clause questions. This main effect did not interact with the status of the *wh*-phrase. The question, now, is what do these results suggest in terms of the grammatical status of island constraints. A proponent of the Hofmeister-Sag approach might say that relative clauses are simply more difficult to process than complement clauses and simple *wh*-phrases are more difficult to process than d-linked *wh*-phrases, and nothing more need be said, i.e., no island-constraint is needed in the grammar. A proponent of an alternative view might say that the inferior status of the relative clause examples is evidence for a grammatical island constraint but the greater acceptability of the d-linked examples in general should be interpreted in terms of processing ease. In short, the results by themselves do not really dictate one approach or the other to the analysis of island constraints.

The results of the experiment do call into question the idea that d-linked interrogatives are easier to process than non-d-linked interrogatives due to memory retrieval of the interrogative phrase or due only to memory retrieval of the interrogative phrase. An obvious alternative to the representational complexity hypothesis is the hypothesis that it is the TYPE of representation that is at issue, not its complexity per se.

As mentioned earlier, Pesetsky (1987) noted that d-linked interrogatives seemed to presuppose that an already salient set was available in context and that a d-linked interrogative picked out a subset of that already established context set. This observation makes it natural to characterize d-linking in terms of the availability of a representation in the discourse representation rather than just in the syntactic representation. This idea fits with other linguistic approaches to filler-gap dependencies, such as Cinque’s (1990) proposal that d-linked interrogatives may enter into either antecedent

government relations or into binding relations characteristic of referential dependencies whereas non-d-linked interrogatives may only enter into the former (see De Vincenzi (1991), for processing evidence supporting the proposal).

In the processing of pronouns, Frazier & Clifton (2002) found that subjects were more likely to choose d-linked antecedents than non-d-linked antecedents in ambiguous sentences like (7) and (8), and reported faster reading times for the d-linked form of disambiguated sentences like those in (9) and (10) in self paced reading.

- (7) a. Rick knew who Janice sang a song to before he went to sleep.
b. Rick knew which visitor Janice sang a song to before he went to sleep.
- (8) a. Who did Brian send a rifle to when he was threatened?
b. Which friend did Brian send a rifle to when he was threatened?
- (9) a. Rick knew who Josh sang a song to | before she went to sleep.
a'. Rita knew who Joan sang a song to | before he went to sleep.
b. Rick knew which sister Josh sang a song to | before she went to sleep.
b'. Rita knew which brother Joan sang a song to | before he went to sleep.
- (10) a. Who did Becky send a rifle to | when he was threatened?
a'. Who did Brian send a rifle to | when she was threatened?
b. Which guy did Becky send a rifle to | when he was threatened?
b'. Which woman did Brian send a rifle to | when she was threatened?

Acceptability ratings also revealed an advantage for d-linked interrogatives in sentences with resumptive pronouns, such as (11a) relative to (11b).

- (11) a. (*) Which students did the teacher wonder if they had gone to the library?
b. * Who did the teacher wonder if they had gone to the library?

They attributed the advantage to the fact that a discourse referent is immediately established for d-linked phrases but not for non-d-linked ones. This hypothesis also explains the results of Frazier et al. (1996), according to which d-linked phrases are immediately instantiated in the discourse representation. In (12c), this means setting up a singular masculine discourse referent when *which boy* is processed. Later, when the embedded clause is processed this decision will need to be revised due to the one-one pairing bias carried by *married*. The assumption that there is just a single boy will need to be abandoned. In sentences like (12) with a universal quantifier scoping over the trace of the interrogative phrase, this should result in difficulty processing the embedded clause of the d-linked example just in case it contains a verb like *marry*. In a self paced reading study, the reading time penalty for (12c) was in fact observed.

- (12) a. Which boy did Tom say that every girl saw?
 b. Who did Tom say that every girl saw?
 c. Which boy did Tom say that every girl married?
 d. Who did Tom say that every girl married?

The preceding considerations suggest that d-linking may confer an advantage, or under some circumstances disadvantage, that is due to qualitative differences in the representation of d-linked and non-d-linked phrases. D-linked interrogatives may immediately receive a discourse representation in addition to their syntactic representation. As a result, the d-linked phrase will be part of a more structured representation. The advantage of being represented in the discourse representation may include persisting longer in cases of memory overload, assuming that discourse representations are more integrated with long-term memory structures than are syntactic representations. But even without memory overload or memory retrieval difficulties, the discourse representation may enjoy an advantage over a syntactic representation, at the very least under circumstances where the dependency in question implies familiarity with the entity in question. This clearly characterizes the use of personal pronouns, which typically have familiar antecedents. But it may also apply normally to listeners who are asked a question about the identity of someone/something. For a question to be felicitous, it must be reasonable for the asker to think that the addressee

is in a position to supply the requested information. So even without the presupposition that comes with a d-linked interrogative, the pragmatics of information-seeking questions may imply that the referent of the interrogative is familiar to the listener. Note that this approach to d-linking can explain why there is an advantage for d-linked interrogatives even in the absence of an explicit contrast set. Fedorenko & Gibson (2009) investigated multiple questions in English and Russian. They found that d-linked interrogatives were rated better than non-d-linked ones even without a discourse context. Like the present work, their results indicate that the advantage for d-linked interrogatives traces not to their accessibility in memory. Rather, Fedorenko and Gibson suggest that *which* + noun maps more transparently onto its set independent of the availability of discourse context. Although they do not develop this idea in detail, at least one way of understanding it is in terms of the discourse instantiation hypothesis defended here.

References

- Chung, Sandra, William A. Ladusaw & James McCloskey (1995). Sluicing and logical form. *Natural Language Semantics* 3, pp. 239–282.
- Cinque, Guglielmo (1990). *Guglielmo Cinque Types of A'-dependencies*. MIT Press, Cambridge, MA.
- Craik, Fergus I. M. & Robert S. Lockhart (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior* 11, pp. 671–684.
- De Vincenzi, Marica (1991). *Syntactic parsing strategies in Italian*. Kluwer Academic Publishers, Dordrecht.
- Enç, Mürvet (1991). The semantics of specificity. *Linguistic Inquiry* 22, pp. 1 – 25.
- Fedorenko, Evelina & Edward Gibson (2009). Syntactic parallelism as an account of superiority effects: Empirical investigations in English and Russian. Ms. MIT.
- Frazier, Lyn & Charles Clifton, Jr. (1998). Comprehension of sluiced constituents. *Language and Cognitive Processes* 13, pp. 499–520.
- Frazier, Lyn & Charles Clifton, Jr. (2002). Processing d-linked phrases. *Journal of Psycholinguistic Research* 29, pp. 125–139.

- Frazier, Lyn & Charles Clifton, Jr. (2005). The syntax-discourse divide: Processing ellipsis. *Syntax* 8, pp. 154–207.
- Frazier, Lyn, Bernadette Plunkett & Charles Clifton, Jr. (1996). Reconstruction and scope. Dickey, Michael & Susanne Turnstall (eds.), *University of Massachusetts Occasional Papers in Linguistics*, Graduate Linguistic Students Association, vol. 19.
- Garnham, Alan, Jane Oakhill, Marie-France Ehrlich & Manuel Carreiras (1995). Representations and processes in the interpretation of pronouns: New evidence from spanish and french. *Journal of Memory and Language* 34, pp. 41–62.
- Hofmeister, Philip (2009a). Representational complexity and memory retrieval in language comprehension. CUNY Poster.
- Hofmeister, Philip (2009b). Retrieval and gradience in filler-gap dependencies. *Proceedings of the 43rd Regional Meeting of the Chicago Linguistics Society*, Chicago Linguistics Society.
- Jonides, John, Richard L. Lewis, Derek Evan Nee, Cindy A. Lustig, Marc G. Berman & Katherine Sledge Moore (2008). The mind and brain of short-term memory. *Annual Review of Psychology* 59:1, pp. 193–224.
- Martin, Andrea & Brian McElree (2008). A content-addressable pointer mechanism underlies comprehension of verb phrase ellipsis. *Journal of Memory and Language* 58, pp. 879–906.
- Merchant, Jason (2001). *The syntax of silence: Sluicing, islands and identifying in ellipsis*. Oxford University Press, Oxford.
- Pesetsky, David (1987). Wh-in-situ: Movement and unselective binding. Reuland, Eric & Alice ter Meulen (eds.), *The representation of indefiniteness*, MIT Press, Cambridge, pp. 98–129.
- Sag, Ivan, Philip Hofmeister & Neal Snider (2009). Processing complexity in subadjacency violations: The complex noun phrase constraint. *Proceedings of the 43d Regional Meeting of the Chicago Linguistics Society*, Chicago Linguistics Society.

Appendix

1. It's possible John witnessed a terrible accident today but he didn't say what.
It's possible John witnessed a terrible accident today but he didn't say which one.
John met someone who witnessed a terrible accident today but he didn't say what.
John met someone who witnessed a terrible accident today but he didn't say which one.
2. It's clear that Brenda heard an awful rumor this week but she didn't say what.
It's clear that Brenda heard an awful rumor this week but she didn't say which one.
Brenda talked to someone who heard an awful rumor this week but she didn't say what.
Brenda talked to someone who heard an awful rumor this week but she didn't say which one.
3. It's obvious that Paula was told about an international scandal but she didn't remember what.
It's obvious that Paula was told about an international scandal but she didn't remember which one.
Paula dates the man who was told about an international scandal but she didn't remember what.
Paula dates the man who was told about an international scandal but she didn't remember which one.
4. It's rumored that Shawn received an unbelievable award but he didn't say what.
It's rumored that Shawn received an unbelievable award but he didn't say which award.
Shawn interviewed the author who received an unbelievable award but he didn't say what.
Shawn interviewed the author who received an unbelievable award but he didn't say which award.

5. It was reported that the photographer won a large prize but he wouldn't divulge what.
It was reported that the photographer won a large prize but he wouldn't divulge which prize.
The photographer filmed the actress who won a large prize but he wouldn't divulge what.
The photographer filmed the actress who won a large prize but he wouldn't divulge which prize.
6. Everyone knows the judge was offered a huge bribe but he wouldn't say what.
Everyone knows the judge was offered a huge bribe but he wouldn't say which bribe.
The judge questioned the witness who was offered a huge bribe but he wouldn't say what.
The judge questioned the witness who was offered a huge bribe but he wouldn't say which bribe.
7. It's apparent that the carpenter used terribly expensive exotic wood but he wouldn't tell us what.
It's apparent that the carpenter used terribly expensive exotic wood but he wouldn't tell us which wood.
The carpenter was jealous of the builder who used terribly expensive exotic wood but he wouldn't tell us what.
The carpenter was jealous of the builder who used terribly expensive exotic wood but he wouldn't tell us which wood.
8. We heard that the professor was honored for an exceptionally original article but then he wouldn't tell us what.
We heard that the professor was honored for an exceptionally original article but then he wouldn't tell us which one.
The professor praised the person who was honored for an exceptionally original article but then he wouldn't tell us what.
The professor praised the person who was honored for an exceptionally original article but then he wouldn't tell us which one.

9. It's clear that Gloria liked some flower near the lotus flowers but then wouldn't tell us what.
It's clear that Gloria liked some flower near the lotus flowers but then wouldn't tell us which flower.
Gloria adored the man who liked some flower near the lotus flowers but then wouldn't tell us what.
Gloria adored the man who liked some flower near the lotus flowers but then wouldn't tell us which flower.
10. We heard that Jason hit an important politician but he wouldn't tell us who.
We heard that Jason hit an important politician but he wouldn't tell us which politician.
Jason knows the guy who hit an important politician but he wouldn't tell us who.
Jason knows the guy who hit an important politician but he wouldn't tell us which politician.
11. We found out that Gretchen invited a famous artist to the party but she wouldn't say who.
We found out that Gretchen invited a famous artist to the party but she wouldn't say which artist.
Gretchen has a boyfriend who invited a famous artist to the party but she wouldn't say who.
Gretchen has a boyfriend who invited a famous artist to the party but she wouldn't say which artist.
12. I know Britney destroyed a new vehicle but she didn't reveal what.
I know Britney destroyed a new vehicle but she didn't reveal which vehicle.
Britney likes this guy who destroyed a new vehicle but she didn't reveal what.
Britney likes this guy who destroyed a new vehicle but she didn't reveal which vehicle.

13. Sam wrote that Carolyn broke some bone but she didn't want to say what.
Sam wrote that Carolyn broke some bone but she didn't want to say which bone.
Carolyn is taking care of some man who broke some bone but she didn't want to say what.
Carolyn is taking care of some man who broke some bone but she didn't want to say which bone.
14. I heard that Charles told an outrageous story to his boss but then he wouldn't tell us what.
I heard that Charles told an outrageous story to his boss but then he wouldn't tell us which story.
Charles likes this girl who told an outrageous story to his boss but then he wouldn't tell us what.
Charles likes this girl who told an outrageous story to his boss but then he wouldn't tell us which story.
15. It's clear that the scientist discovered an important cure but he refused to say what.
It's clear that the scientist discovered an important cure but he refused to say which cure.
The scientist met a doctor who discovered an important cure but he refused to say what.
The scientist met a doctor who discovered an important cure but he refused to say which cure.
16. It's not surprising that the diplomat heard about a new coup attempt but he wasn't saying what.
It's not surprising that the diplomat heard about a new coup attempt but he wasn't saying which attempt.
The diplomat talked with a secretary who heard about a new coup attempt but he wasn't saying what.
The diplomat talked with a secretary who heard about a new coup attempt but he wasn't saying which attempt.

From nounphrase ellipsis to verbphrase ellipsis: The acquisition path from context to abstract reconstruction*

Tilbe Goksun¹, Tom Roeper², Kathy Hirsh-Pasek³, and Roberta M. Golinkoff⁴

¹*Temple University*

²*University of Massachusetts*

³*Temple University*

⁴*University of Delaware*

1. Introduction

The rules of grammar allow us to reconstruct arguments even from limited linguistic form. Such is the case of ellipsis. This very common device of language requires adults and children to exercise pragmatic inference by connecting what they hear either to discourse or context. A sentence like “put it there” requires us to compute where “there” is. Almost every sentence or dialogue contains elliptical reference. If you ask “do you want that” and get the answer “yes”, it must mean “yes I want that” not “yes I’m going outside”. This paper investigates whether children rely on pragmatic inference or language discourse to interpret elliptical utterances. The analyses herein, and the new data presented, suggest that children move from a pragmatic approach to a linguistic one during the course of development. If we say: “here’s cake, want some” the child might fix “some” the way he determines the meaning of “there” by just looking around, while the adult does it by sticking in cake after some to get some cake. Put differently, suppose the child understands “want something” for “want some” and

*This work was supported by an NICHD grant (5R01HD050199) and by an NSF grant (BCS-0642529) to the last two authors. Portions of this research have been presented at the Conference on Human Development (2006) and at the Biennial Meeting of the Society for Research in Child Development (2007). We thank everyone at Temple University Infant Lab and University of Delaware’s Infant Language Project for their invaluable contributions to this project. Special thanks to Meredith C. Jones for helping the data collection.

guesses that the topic is the cake she sees and says “yes.” She would have the right answer obtained through the wrong means. These two totally different strategies divide the world’s languages. In Huang (1982)’s terms the “hot” languages refer to context, where “cool” languages require reference back to discourse. These are broad terms that refer to how easily a particular language allows reference to context. However, it has very specific grammatical consequences: arguments are deletable in “hot” languages where they are required in “cool” languages. A simple example, often found as an L2 error, is the sequence: “here’s a donut. Do you want?” In English one must say “do you want it”, but not in Chinese. It is possible that children begin by assuming that all languages are “hot” and therefore context-dependent, which is why many acquisition researchers have intuitively asserted early context-dependence.

Adult lack of awareness about ellipsis often leads to discourse gridlock with children, still struggling with the intricacies of reconstruction. Consider this dialogue (Roeper 2007):

- (1) Mother: Do you want some milk or do you want some juice?
 Child: I milk juice [?]
 Mother: Huh?
 Child: Milk juice.
 Mother: No, you can either have one or the other. You can(’t) have both.
 Child: Milk juice.

The child can obviously not deal with *one or the other, but not both*, which must be reconstructed as “one of the milk or juice or the other of the milk or juice but not both the milk and the juice.” This seems like quite a task for the child.

1.1 Child Ellipsis

Jensen & Thornton (2007) report examples of successful production from Nina (2; 3 years of age) of some NP-ellipses in cases like:

- (2) Mother: Whose hat is that?
 Nina: Mrs. Wood’s

In this case, the *hat* is implied after “Mrs. Wood’s __”. But, full control of ellipsis is much more complex as Wijnen et al. (2003) have shown

with Dutch children some of whom reconstruct the argument but not the adjunct in:

- (3) Three girls are in the sandbox. Are two __ upsidedown?

Scene: two girls *outside the sandbox* are upsidedown.

Adults reply “no” because they reconstruct “are two [girls in the sandbox] upsidedown”, while 4-year-olds often say “yes,” because they reconstruct only: are *two [girls] upsidedown*. Both 3-year-old Dutch and 4-year-old English-speaking children said “yes” more than 80% of the time. When children were presented with a similar scene in which children were standing in a normal position in the sandbox and two *adults* were standing upsidedown *outside* the sandbox, their “yes” answers appropriately decreased to 36% (3-year-old Dutch children) and 27% (4-year-old English children). However 36% “yes” means that even some younger children will allow a free contextual reference to the object [two *free pronoun* \Rightarrow *adults*], if two adults are upsidedown anywhere. This is the first evidence that Nounphrase Ellipsis (NPE) can be challenging for children, especially 3-year-olds. The evidence below directly supports this.

1.2 Adult Ellipsis

To avoid the sin of understating the child’s ultimate task, let us consider what kind of knowledge of ellipsis the adult has. First, languages do not always allow the same forms of ellipsis. In Japanese one can say:

- (4) I want a hat or I want to go and Bill wants
 \Rightarrow [a hat or to go].

In English we have to add either an object ‘*one*’ or ‘*to*’:

- (5) I want *one* or I want *to*

This difference will prove important below. Virtually no other language allows Verbphrase Ellipsis (VPE) with just dangling *to*. These distinctions immediately imply that there will be an acquisition path that requires triggering evidence for the child to see each of them.

In general, there are three ingredients that can actually interact: 1) pragmatic reference (above), 2) syntactic reconstruction (copying what is said) and 3) an abstract semantic form which captures variable behavior (below). We could paraphrase these options as:

- (6) a. look at context, or
- b. copy exactly what occurred, or
- c. project a meaning, Logical Form, which captures hidden relations, like “bound variables.” Bound-variables occur in many places, but a simple example is with quantifiers like: *every boy lost his hat*, where every/his are pairwise bound (= each boy lost a different hat, his own).

1.3 VP-ellipsis in Acquisition

If we take a cross-linguistic perspective, we find that NPE is by far more common than VPE in the world’s languages and therefore by classical typological reasoning, it should be more available to children. In particular, the varieties of VPE in English, particularly the hanging *to* (e.g., I want to) are virtually unknown elsewhere and should prove a challenge to children. We will suggest below that an efficient *Interface Preference* may be allied with, or in fact, explain this preference.

VPE has received the most attention in both linguistic theory and acquisition (e.g., Foley et al. 2003; Lust & Foley 2004; Postman et al. 1998; Thornton & Wexler 1999). It involves the reconstruction of an entire VP, including Verb, object, and modifiers:

- (7) Every man painted his car carelessly, and so did every boy
 ⇒ [paint his car carelessly]

There are two properties of special interest here: the pronoun *his* and the adjunct *carelessly*. The adjunct *carelessly* can be either included, as above, or substituted for: [and so did every boy __ *carefully*]. *His* allows *sloppy-identity*, which means it may or may not switch reference from every *man’s* car to every *boy’s* car. A crucial feature is that the antecedents (every man, boy) are higher in the tree (c-commanding) than their possessives (*his*). In addition, each link involves an entire set of people (men, boys). It is commonly represented with a semantic device called *lambda abstraction* (which pulls out a common property from two clauses into an abstract structure):

- (8) Oscar bites his Banana and Bert does __ too.

Classic derivation of the sloppy reading (Williams 1977):

- (9) a. Derived VP rule:
 Oscar $[_{VP} \lambda x(x \text{ bites his banana})]$ and Bert does $[_{VP}[_{V} e][_{{NP}}[_{N} e]]]$ too
- b. Variable rewriting rule:
 Oscar $[_{VP} \lambda x(x \text{ bites his banana})]$ and Bert does $[_{VP}[_{V} e][_{{NP}}[_{N} e]]]$ too
- c. VP rule (= VP copy): Oscar $[_{VP} \lambda x(x \text{ bites his banana})]$ and Bert does $[_{VP} \lambda x(x \text{ bites his banana})]$ too

It is this bound-variable property which some beautiful experiments by Barbara Lust have investigated with children, showing apparently, that they control the *variable-binding* entailed at Logical Form. Every other semantic theory entails some kind of additional abstract level of representation (see discussion in Foley et al. (2003)) where they also propose a syntactic instead of a semantic analysis for VPE. Foley et al. (2003) claimed that children have correct VPE, which makes these observations surprising: “how is it that competence for both types of interpretation (sloppy and strict) appears to be acquired so early and simultaneously, with no developmental lags between them? Why would this be so if the types of interpretation involve qualitatively different representations and derivations, as suggested in previous proposals? For example, if differentiation of pronoun types in syntax (Fiengo & May 1994) is necessary to generate the two types of interpretation, what predicts early and simultaneous access of the two pronoun types? Why, given earlier representations, would the sloppy interpretation be so strongly preferred, the strict interpretation so much less frequent, in spite of the grammatical availability of both? (p.71)” Significantly, there is a contrast: quantificational sentences involve such an extra layer of semantic structure, while bare object cases do not:

- (10) I want some ...

This sentence can be resolved, from this perspective, by pure copying without that extra level, which is why NPE is radically different from VPE. However for Japanese it has been argued that NPE is a subpart of VPE, which we show below¹.

¹ These claims remain controversial (see Johnson 2008; Tomioka 2008; Hardt 1992, 1999, 2008). Our goal is to contribute to this discussion by showing that acquisition evidence and a simple in-

This will develop into the core of our hypothesis: NPE is inherently simpler for the child and therefore precedes VPE. Once we have laid out further facts about English, here is where the analysis is leading: If the VPE cases are related to the NPE cases, then, apparent bound-variable interpretations may not require the extra layer of structure. In addition, children may *convert* VPE that they hear into NPE, which has consequences for the experiment below. We argue that there is a path indicating that they do not have variable-binding at first, simply apparent variable-binding, and that their predictions ought to be a more intricate path. Their syntactic proposal moves in this direction as well, but does not include the crucial possibility of free pronominal reference.

2. Connections to Quantification in Acquisition

This approach, detailed below, would fit other evidence that quantificational aspects of meaning are acquired later (e.g., Roeper et al. 2006). In a word, children acquire collective “all” before the age of 2 years, but real quantifier-variable structures such as “every N” (*every boy has a hat*) do not appear reliably until 4 to 5 years of age. It means that 2- to 3-year-old children, at least for ellipsis, may not yet engage higher order semantic structure. These data would be consistent with the idea that NPE may precede VPE in acquisition.

By “higher order semantic structure,” we refer to a family of semantic theories which express the quantificational relation with either an entire layer of structure or an extra operation of abstraction or extraction. All the theories argue for something extra in the semantics.

terface principle argue in behalf of the view that a simple PRO form can be involved. We do not assume that the current formulations for adult grammars will prove perfect for the representation of child grammars. This may lead to the claim that the adult grammars should be revised to make a connection to the acquisition path more explicit. In other words, our goal is to formulate the connection between VPE and NPE for children in its own terms. In particular, it can be argued that the same semantic reconstruction is involved in both NP and VP ellipsis. However we argue that the acquisition path favors a theory which makes a fundamental distinction between pronominal reference to context and variable semantic reconstruction. Another alternative is to argue that a child version of NPE exists with simple copying, no distinct LF, and then it is replaced by a more sophisticated form. (Thanks to Kyle Johnson for help in clarifying these and other points.)

2.1 Theoretical Background

In an excellent introduction, Johnson (2008) alerts us to the fact that variable-binding might be illusory, a perhaps crucial clue to envisioning the acquisition path. We need to walk through the examples to appreciate their force (Evans 1977):

- (11) When John cooks something, he won't acknowledge what he can't, and when he bakes something, he won't either.²

This has to be reconstructed as: *won't* acknowledge what he can't [bake] *either*. This reconstruction is an exact copy of the prior VP, except for the last verb (cook) which gets a substitution (bake). This requires a hidden pronominal form for the verb, so we can substitute *bake* for *cook*. It is not a bound-variable substitution like with "his" and lacks a c-commanding *verb* higher in the tree. It can be argued that the same kind of operation is at work with other cases where the c-command requirement is not met:

- (12) The police officer who arrested John insulted him and the one who arrested Bill did too.
[\Rightarrow insult him = Bill]

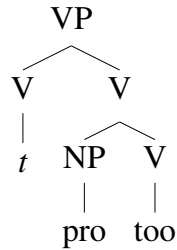
Here the crucial antecedents *John* and *Bill* are buried in relative clauses so they fail the c-command requirement, of an NP higher in the tree, and yet we still have a seeming "bound variable" effect:

- (13) "the pronoun *him* contains a hidden description that makes it equivalent to, say, the expression *that guy* and this creates the illusion that there is a bound pronoun in the ellipsis = the police officer who arrested John insulted [him - that guy who got arrested]."

What may seem like an arcane example opens an important acquisition option: could the child, as we have hinted, have a simple free pronoun with definite reference instead of a bound-variable? This would make a combination of simple copying plus a pragmatic reference to context a plausible first stage for the child, without an LF form of Bound variables, but a *trace* and an empty pronoun instead:

² Example discovered by Dan Hardt and Bernard Schwartz independently.

- (14) John washed his car and Bill did *v* [*wash* [*pronoun*]] too.



A *trace* requires copying the exact word *wash*, and a true pronoun like (did wash *it* too) would get identical reference (John's car).

However, the empty *pro* also has the potential to pick out a *new* object from context. Unlike a *trace*, it does not copy its reference from the higher clause. Thus it would not be bound to discourse for children, who could use context, although it would be discourse-bound for adults. This means that it could get an extra meaning beyond the other noun: it could mean Bill washed *Fred's* car. To put this in very simple terms, fixing the pronoun reference should require nothing more semantically complex than fixing deictic reference if one says "wants that".

Theoretical claims have some bite when they make specific predictions:

- (15) Claim: children project a simple missing object for VPE and NPE in early stages.

This leads to an experimental prediction (see Roeper 2007), which in fact has been borne out: children might have too free a reference here, the definite reference could be to the higher NP *or* to context itself, which is where we intuitively believe that children start. Matsuo (2007) and Vasic et al. (2007) showed that Japanese and Dutch children allowed exactly a *third reference* for:

- (16) John washed his car, and so did Bill.

where Bill washed *Fred's* car, quite at odds with adult English (but not Japanese)³ This goes beyond what adults allow, but it suggests that the

³ Matsuo used examples like:

- (17) Kuma-san-ga aoi osakana-o mitukemasita.
bear-Mr.-NOM blue fish-ACC find-PAST

apparent bound-variable reading, justifying the LF, may have a different origin that does not require the LF, but requires greater leaning on context, which acquisition theorists have often intuitively assumed.

2.2 Japanese Ellipsis

In fact, recent theoretical work in Japanese (Hoji 1998; Otani & Whitman 1991; Sugisaki 2007) has proposed precisely that VPE has NPE inside of it and it is linked to the kind of construction not found in English:

- (19) John wants a car and Fred wants __ too.

That is, in effect, Japanese extracts the noun part of the VPE and allows it to be a pronoun, possibly filled in by context (predictable as a “hot” language), and notably, does not carry the adjunct form with it. Thus the Japanese equal to this sentence:

- (20) John can paint a car carefully, but Bill can’t.

can’t get the reading *can’t paint a car* instead of a quite different *can’t paint a car carefully* which we find in English where we still presuppose that a car gets painted. We have in effect:

- (21) Bill didn’t paint [NP = free pronoun]

and the adverb remainder (carefully) of the VP may not reconstruct at all (if child English is like Japanese, then we should predict the same non-inclusion of adverbs by English children, suggesting an obvious experiment.)

Saito (2003) has provided arguments, based on sensitivity to subadjacency, that this empty category is not a pro, but a form of *argument deletion*.

‘The bear found a blue fish,

- (18) Sosite, tora-san-mo [NOSAKANA-O] mitukemasita.
 And tiger-Mr.-ALSO fish-ACC find-PAST
 and the tiger did [_{VP}find a fish], too’

where the tiger found a pink fish, not a blue one, but this is acceptable. From her data, she relegates this to a minor possibility for younger children, but that is exactly the stage we are seeking to capture.

For our purposes, the important point is that reference to context is required to fill the object rather than a form of sloppy identity⁴.

- (22) John bir arabayı dikkatlice boyuyor, fakat Bill boyayamıyor.
 John a car-ACC carefully pat, but Bill can't-paint
 'John paint the car carefully, but Bill can't paint ----'

Turkish does not allow using "can't" without the verb. However, one can still read the sentence similar to Japanese either *can't paint a car* or *can't paint a car carefully*.

Adult English, however, has the classic form of NPE which is simply the deletion of an object:

- (23) John has bananas. Bill wants some [bananas]

It still does not allow the bare **and John wants [free Obj]*. In contrast, the use of bare verb "want" is possible in Turkish.

- (24) John'ın muları var. Bill de istiyor [muz].

'John has bananas. Bill wants [bananas] too.'

In Turkish, it is possible to use "want" without an NP. In non-elliptical sentences, one might say either "Bill wants some bananas" or "Bill wants bananas."

Now we need to take a closer look at NPE.

2.3 Nounphrase Ellipsis

How does NP-ellipsis really work? NPE, while simpler than VPE, also involves, for adults, some licensing relations which vary across languages.⁵ We do not allow NPE after adjectives:

- (25) John has a big blue house, **and John has a small red*.

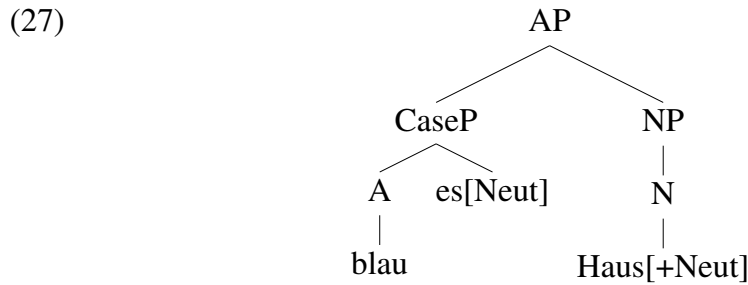
However, Germanic languages do because they have an agreement element that is morphologically marked:

⁴ This interesting claim raises questions for the theory of empty categories and increases the set of options to which a child might be sensitive. Other contexts exist where argument-deletion might apply, but the challenge is then to prevent the child from overgeneralizing them.

⁵ See Lobeck (1995) for a comprehensive discussion of NPE.

- (26) John hat ein blaues Haus und Fritz ein kleines rotes...

Here, the adjective *rot* (red) carries a neuter marker (-es), which agrees with the object noun *Haus* (neuter *house*).



The two [+Neut] markers are in Agreement and the Agreement relation then licenses the deletion of the Noun. This analysis entails the presence of higher structure in the Determiner Phrase, which being particular to some languages, requires acquisition (see Wijnen et al. 2003).

English has such higher structure for Numerals as well:

- (28) John has a hat, and Bill has five ...

Note that what must be provided to license an empty object is the plural: has (five [+pl] (hat)[+pl]).

So English has an invisible form of plural engaging the same kind of Agreement found in German. Evidence for this view comes from work on *one* by Sugisaki (2005).

2.4 Naturalistic data on *one* ellipsis

Sugisaki provides extensive acquisition evidence that the Numeral licenser is present for very young children when they use *one* (only one exception in CHILDES), children never say **five ones*, but easily say *blue ones* because the adjective without a plural marker in English does not license the empty object, but the numeral does (“I have five”). He uses this as an argument against the idea that *one* is just a pronoun as is often argued when children say: “I want one”. Sugisaki⁶ shows that *one* should occupy the same Numeral position in a DP as when they say I want *two*, which obviously

⁶ Lombert-Huesca (2002) in Sugisaki: “one is inserted as a Last-Resort operation when it is necessary to give phonological support to the Number affix that would be stranded otherwise”

licenses an empty N. With an extensive study of early production, Sugisaki argues that children always use bare Numerals to license missing objects. Children do not say **five ones*, it follows that *one* is excluded because it occupies the same Numeral position prior to the Noun as five. If *one* is like *five*, then it also licenses an empty object elliptically: one [N] and five [N].

3. Our Experiment: Preferential Looking Paradigm and Ellipsis

Now, where does our experiment fit in? Our goal is to take the first steps in contrasting VPE and NPE, and to explore methods that might allow experimentation, especially anti-pragmatic experiments with young children where we might see what a child's initial assumptions are about ellipsis.

3.1 The Initial State

Our somewhat intricate analysis leads to a vision of the child's initial state, the Default assumptions about how ellipsis should work. While the earliest stages may be a kind of inferential "guess" based on everything available, what does the child do when she a) recognizes that something structural is missing, and b) seeks to model it on the immediately previous utterance? This step itself, logically, requires an assumption, namely that it is precisely the immediately preceding utterance from which information must be taken. There is another option: take the information from *Context*.

This idea needs to be more carefully situated. In particular, it is an aspect of the interface between grammar and pragmatics. One natural, but not necessary, hypothesis is that one expects *modular homogeneity* to be preferred to any representation that seems to involve an interface with other mental abilities. Consider the idea intuitively: A phonological analysis should be easier than a simultaneous phonological and syntactic analysis. Coordinating your hands and feet should be harder than just coordinating your hands. However, biology offers the opposite as well: instant, well-defined interactions between different organs, like the heart and the lung. Suppose we say that grammar, particularly acquisition, favors a one-step connection between syntax and context. That is the comprehension system is most efficient when it can project a reference for a sentence with minimal further syntactic and semantic operations. Let us state this informally.

3.2 Interface Principle

Our argument is that the child seeks an immediate connection between sentences and context with a minimum of extra mental operations on the linguistic object.

- (29) *Interface Principle*: there is a one-step connection between syntactic representation and contextual interpretation.

Ultimately, the adult will prefer a within-language interpretation of ellipsis over a contextual one: connect an ellipsis site to previous verbal discourse rather than visual context. However, pressure toward efficient referential interpretation makes this Interface Principle plausible.⁷

This intuitive statement is programmatic and far from adequate because the presupposed syntax or semantics may still be non-obvious in making the contextual connection. For instance, if I say: “that” as compared to “drink that,” *that* is not interpreted alone, but requires the verb *drink* to choose the right object. Nevertheless, this principle can be a useful starting point. Our argument now leads to this prediction, if NPE involves use of a pronoun which enables this efficient interface, then NPE should be easier and then, possibly, if Japanese or Turkish is the model, the child may use NPE as a part of VPE to reconstruct what is missing.

3.3 Acquisition Path

So now we can formulate an acquisition path for ellipsis:

1. Free unpredictable inference on context = guess what the parent means.
2. Correct NPE: project missing object from previous sentence or context.
3. Incorrect VPE Reconstruct verb from the previous sentence by copying, but project open empty object [=Japanese Nounphrase-ellipsis]
4. Adult stage: VPE copy full VP in ellipsis site, project bound variable structure in semantics.

⁷ See Hardt (2008) for arguments that discourse representations can carry over sloppy-identity readings. This argument also enlarges the interpretation of ellipsis beyond the c-command domain, just as our use of Context does.

- a) Disallow free reference to context.
- b) Allow recursive embedded structures to reconstruct.

Our experiment is consistent with this projected path, but covers only a portion of it. The experiment sets up either the verb or the object as the point of contrast and provides the child with both. Our experiment involved minimal pair [want to/want one] and identical events with a Preferential Looking Paradigm (PLP) using pointing as the dependent variable (Golinkoff et al. 1987; Hirsh-Pasek & Golinkoff 2006):

Children and adults could indicate their interpretation of the elliptical sentences by pointing to a screen. Each of the two sides of the screen is consistent with a different interpretation of the linguistic input. Participants were presented 6 video clips (two introductory and four test trials) involving live actions performed by adult actors on a large television. Importantly, the video clips were identical for both noun and verb ellipsis conditions; only the auditory stimulus varied. The pivotal device here is contrasting visual situations where we have different verbs with the same noun or different nouns with the same verb. In the picture Ali eats a banana and the child hears, “Ali is eating a fruit. Look she is eating fruit.” Then, the child hears, “Now Hannah wants one” or “Now Hannah wants to.”

- (30) a. wants one \Rightarrow one banana.
- b. wants to \Rightarrow to eat fruit.

(30a) should lead to a picture where Hannah holds or waves a banana. (30b) should lead to a picture where Hannah eats a fruit (= a different fruit, an orange).

If the child takes (a) for both, then they are giving an NPE response to both NPE and to VPE contexts. That is, they convert *wants to* into just *wants* with an empty noun after it (want empty pronoun] instead of a missing verb complement (to [eat fruit]) The Appendix shows the full list of stimuli. Two introductory trials familiarized children with the testing procedure and the fact that they would be asked to point. This was followed by four blocks of test trials comprised of a video of the setting event (12 sec) and a pair of split-screen test events (12 sec). In between each trial, children’s attention was heightened by introducing a 3-second inter-trial interval showing a giggling baby face in the center of the screen.

Now we review the exact sequence. In the introductory phase, children saw a truck and a doll on the split-screen for 6 seconds and from the speaker they heard ‘Look! Can you find the truck? Where is the truck?’ The second introductory clip was of a man dancing for 6 seconds on the full screen after which the same man was seen dancing on one side of the split-screen and drinking on the other side for 12 seconds. Children heard ‘Can you find dancing? Where is he dancing? Point to dancing!’ At test, children saw a full screen of an adult doing an act. For example, they were presented an adult eating a banana. After the intertrial interval, children were shown a pair of events on a split-screen. On one side, the girl was eating an orange (same verb-different noun) while on the other side, she was waving a banana (same noun-different verb). In Nounphrase Ellipsis condition, while watching the full screen children heard ‘Alli is eating a banana.’ At split-screen phase, they heard, ‘Point to the picture where Hannah wants *one*.’ On the trials that have mass nouns, children were asked to ‘Point to the picture where Hannah wants *some*.’ Similarly, at Verbphrase Ellipsis condition, while watching the adult eating a banana, children heard ‘Alli is eating a fruit,’ after which they heard ‘Point to the picture where Hannah wants *to*’ in split-screen. The use of the term ‘fruit’ allowed us to offer two fruit choices so that “eat fruit” would refer only to the act of eating. Only the first pointing the child did was accepted as the child’s response.

3.4 Results

The results suggest that even adults found our elliptical tasks challenging. Only 83% and 88% of the correct responses were given for Nounphrase and Verbphrase ellipses, respectively. While these levels are well above chance, they suggest that reconstructing elliptical sentences with the minimal contexts provided here is a difficult task. Yet, adults are clearly in command of the grammatical markers we selected. Both 3- and 4-year-old children showed some sensitivity to elliptical sentences. This fact is interesting in and of itself given how infrequently children hear precisely these structures in the input (although most dialogues contain some kind of ellipsis). Four-year-olds were responded at adult levels for both Nounphrase and Verbphrase ellipsis (77% vs. 75 %, respectively). Three-year-olds, in contrast, were only successful in Nounphrase ellipsis (76% vs. 56%, NPE and VPE, respectively) where 50% is a chance result. This suggests that devel-

Age group	Nounphrase Ellipsis Condition “one” - 3 trials				Verbphrase Ellipsis Condition “to” - 4 trials				
	0 correct	1 correct	2 correct	3 correct	0 correct	1 correct	2 correct	3 correct	4 correct
3-year-olds (n = 16)	1	1	6	8	2	6	3	3	2
4-year-olds (n = 16)	2	2	2	10	1	0	3	5	7
Adults (n = 10)	0	2	1	7	0	0	1	3	6

Table 1: The number of correct responses (max = 3 and min = 0 and max = 4 and min = 0, in nominal ellipsis and verbal ellipsis conditions, respectively) by children and adults.

opmentally, NPE is slightly easier than verbal ellipsis. For example, when they heard ‘Alli is eating an apple and Hannah wants one’, they inferred the meaning that one refers to ‘an apple.’ Similarly, when they heard ‘Alli is eating a fruit and Hannah wants *to*,’ 4-year-olds, but not 3-year-olds interpreted the “*to*” as referring to the act of eating.

That is, if we look at Table 1, we find that

- (31) $5 \Rightarrow 3$ or 4 correct, which means they understand VPE
 $3 \Rightarrow$ chance because they get half of them right
 $8 \Rightarrow 0$ or 1, which means from 75-100% of the time children point to the form with an identical noun (banana). This group among the youngest group is not operating by chance, but choosing NPE analysis instead of a VPE analysis. Their grammar is exactly what we would predict if the child VPE reconstruction contains a structure like:
- (32) wants to = V [pronoun] = want [free pronoun] where eating is ignored, but the empty pronoun is linked to context and NPE so that it is like want one and the identical banana noun is chosen rather than the adult:
- (33) wants to = [eat fruit]VP where a person reconstructs eat fruit which allows any object in the picture, like *orange*.⁸

⁸ The (32) choice actually would allow a one [anything] reading, but that was not made available in

Our analysis includes an important child alteration. In order to convert a VPE into an NPE, the child would have to ignore *to* - the most unusual part of English- that is to ignore the Inflectional Phrase (IP) link indicated by *to* (i.e., want to \Rightarrow want). If they do so, they would have exactly the Japanese form:

- (34) John wants a hat, and Bill wants ___ too.

Preliminary results from Turkish replication of the study indicated that both 3- and 4-year-olds found these elliptical sentences ambiguous. Both NPE and VPE might be interpreted in different ways:

- (35) Alli muz yiyor. Hanna bir tane istiyor.
Alli banana eating. Hanna one wanting.
'Alli eats a banana. Hanna wants one.'

"One" in Turkish sentence usually refers back to the object 'banana.' However, a slightly possible option is that it might signify 'eating a banana.' So, when the child could not find 'eating a banana,' but eating an apple, she might point at the eating action. The VPE is exactly the same as the Japanese case:

- (36) Alli muz yiyor. Hanna istiyor.
Alli banana eating. Hanna wanting.
'Alli eats a banana. Hanna wants.'

Turkish children would go to either NP or VP ellipsis in this kind of sentence; because there is no IP indicating that it is the verb that is elided from the sentence. This is very similar to the recent theoretical work in Japanese (Hoji 1998; Otani & Whitman 1991; Sugisaki 2007), which claims that VP ellipsis has NP ellipsis in it. Thus, the Turkish sentence without "to" might indicate the dropping of the object 'banana', the verb 'want' or the VP 'want a banana.' In our approach, children should prefer a simple missing object and refer to the context to decide what object is missing.

Evidence from Jensen & Thornton (2007) reveal that children avoid this ellipsis of IP. They report that children do not answer wh-questions like:

our experiment, but would be expected for young children given similar results from Wijnen et al. (2003) where [three girls \Rightarrow and two [adults] upsidedown] for a group of younger children.

- (37) “Who ate the ice cream?” with “Dad did” but tend to use full sentences or more than is necessary:
 Mother: Who did you feed?
 Nina: Feed the llama (T3, 1;11)
 Mother: What is the little girl holding?
 Nina: Holding a flower (T3, 1;11)

In addition, Foley et al. (2003) show repetition evidence that young children will leave out exactly the IP information:

- (38) Model: Grover licks his ice-cream and Big Bird does too. Child: Grover lick his ice-cream and Big Bird too (age 2; 11, Foley et al. (submitted))

The child repeats both clauses but drops the IP information [-s, does] from them. This is exactly equivalent to dropping *to* from *want to* thereby converting an English VPR ellipsis into a kind of Japanese NP-ellipsis. The eight young children appear to do precisely that: they do not choose the verb, rather the identical noun screen, just as they do for NPE cases (*want one*).

4. Conclusions

In conclusion, our experiment points the way toward seeing an intricate acquisition path for ellipsis which begins with a strong contextual assumption about reconstructing missing NP's, then applies that analysis to IP + VP reducing it to a Japanese variety VPE, where the object is interpreted by NPE. Finally, as is to be expected if VPE involves higher order quantification, the child comes to understand that bound-variable structure is allowed, which permits a variable understanding of pronouns like *his* but excludes a completely free reference to context for understanding the missing NP. The argument is stronger because we have drawn common conclusions from radically independent means of linguistic analysis: cross-linguistic work, a variety of experiments, naturalistic data, and intuition-based theoretical reasoning. These analyses collectively re-inforce each other while none alone provides a full analysis. This is the deepest and strongest kind of scientific reasoning. The Preferential Looking Paradigm ultimately offers a chance to look at the very earliest forms of elliptical understanding without a layer of

conscious reasoning. Eye-movements, even less connected to the unknown impact of consciousness than pointing, may be able to re-inforce this conclusion with even younger children. This work expands the arguments of Hirsh-Pasek & Golinkoff (2006) about the different status of Nouns and Verbs in the acquisition process. In particular, it argues for a contextual available for Nounphrases as opposed to Verbphrases when examined with the structural diagnostic tool of ellipsis. In addition, it has led to a simple but intuitive claim, the Interface Preference Principle, about how and when a child uses Context to interpret sentences. It has always been obvious that Context serves to validate grammatical interpretations. We argue that it plays a critical role at early stages in giving interpretations to sentences with logical gaps that are a step ahead of where a child's grammar is.

References

- Evans, Gareth (1977). Pronouns, quantifiers, and relative clauses. *Canadian Journal of Philosophy* 7, pp. 467–536.
- Fiengo, Robert & Robert May (1994). *Indices and Identity*. MIT Press, Cambridge.
- Foley, Claire, Zelmira Núñez del Prado, Isabella Barbier & Barbara Lust (2003). Knowledge of variable binding in vp ellipsis: Language acquisition research and theory converge. *Syntax* 6, pp. 52–83.
- Foley, Claire, Julie Pactovis & Barbara Lust (submitted). Links between LF and PF: New evidence from first language acquisition of VP ellipsis .
- Golinkoff, Roberta Michnick, Kathryn Hirsh-Pasek, Kathleen M. Cauley & Laura Gordon (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language* 14, pp. 23–45.
- Hardt, Daniel (1992). A semantic theory of verb phrase ellipsis. *Proceedings of the Second Conference on Semantics and Linguistic Theory*, Columbus, Ohio.
- Hardt, Daniel (1999). Dynamic interpretation of verb phrase ellipsis. *Linguistics and Philosophy* 22:2, pp. 187–221.
- Hardt, Daniel (2008). Vp ellipsis and constraints on interpretation. John-

- son, Kyle (ed.), *Topics in Ellipsis*, Cambridge University Press, Cambridge, pp. 15–30.
- Hirsh-Pasek, Kathy & Roberta Michnik Golinkoff (2006). *Action meets words*. Oxford University Press, New York.
- Hoji, Haime (1998). Null object and sloppy identity in Japanese. *Linguistic Inquiry* 29, pp. 127–152.
- Huang, Cheng-Teh James (1982). *Logical relations in Chinese and the theory of grammar*. Ph.D. thesis, MIT, Cambridge, MA.
- Jensen, Britta & Rosalind Thornton (2007). Fragments of child grammar. Paper presented at the Linguistics Association of Great Britain, King's College London, UK.
- Johnson, Kyle (ed.) (2008). *Topics in Ellipsis*. Cambridge University Press, Cambridge.
- Lobeck, Anne (1995). *Ellipsis: Functional Heads, Licensing and Identification*. Oxford University Press.
- Lust, Barbara & Claire Foley (eds.) (2004). *Language Acquisition: The Essential Readings*. Blackwell.
- Matsuo, Ayumi (2007). Differing interpretations of empty categories in English and Japanese VP-ellipsis contexts. *Language Acquisition* 14, pp. 3–29.
- Otani, Kazuyo & John Whitman (1991). V-raising and VP-ellipsis. *Linguistic Inquiry* 22, pp. 345–358.
- Postman, Whitney, Claire Foley, Lynn Santerlmann & Barbara Lust (1998). Evidence for strong continuity: New experimental results of children's acquisition of VP ellipsis and bound variable structures. *MIT Working Papers in Linguistics: Proceedings of the 8th Student Conference in Linguistics*, MITWPL.
- Roeper, Tom (2007). *The Prism of Grammar: How Child Language Illuminates Humanism*. Bradford Books.
- Roeper, Tom, Uri Strauss & Barbara Zurer Pearson (2006). The acquisition path of the determiner quantifier every: Two kinds of spreading. Heizmann, Tanja (ed.), *UMOP 34: Papers in Language Acquisition*, GLSA, Amherst MA.
- Saito, Mamoru (2003). A derivational approach to the interpretation of scrambling chains. *Lingua* 113, pp. 481–518.

- Sugisaki, Koji (2005). "one" issue in acquisition. Otsu, Y. (ed.), *Proceedings of the Sixth Tokyo Conference on Psycholinguistics*, Hituzi Syobo, Tokyo.
- Sugisaki, Koji (2007). The configurationality parameter in the Minimalist Program: A view from child Japanese. Caunt-Nulton, H., S. Kulatilake & I. Woo (eds.), *Proceedings of the 31st Annual Boston University Conference on Language Development*, Cascadilla Press, Somerville, MA, pp. 597–608.
- Thornton, Rosalind & Kenneth Wexler (1999). *Principle B, VP ellipsis and Interpretation in Child Grammar*. The MIT Press, Cambridge, MA.
- Tomioka, Satoshi (2008). A step-by-step guide to ellipsis resolution. Johnson, Kyle (ed.), *Topics in Ellipsis*, Cambridge University Press, Cambridge, pp. 210–229.
- Vasic, Nada Vasić, Sergey Avrutin & Esther Ruigendijk (2007). The economy hierarchy of referential dependencies in child language: evidence from Dutch VP-ellipsis. *Proceedings of GALANA*.
- Wijnen, Frank, Tom Roeper & Hiske van der Meulen (2003). Discourse binding: Does it begin with nominal ellipsis? *Proceedings of GALA, LOT*, Utrecht.
- Williams, Edwin S. (1977). Discourse and logical form. *Linguistic Inquiry* 8, pp. 101–139.

Video and audio stimuli for both nominal and verbal ellipsis test trial blocks.

	Video	Nominal Condition Audio	Verbal Condition Audio
Test 1	Full screen: Alli eating a banana	Alli is eating a banana. Look, Alli is eating a banana! Hey, Alli is eating a banana.	Alli is eating a fruit. Look, Alli is eating a fruit! Hey, Alli is eating a fruit.
	Split screen: Hannah eating orange on; Hannah waving banana	Point to the picture where Hannah wants one! Show me the picture where Hannah wants one. Point to where Hannah wants one!	Point to the picture where Hannah wants to! Show me the picture where Hannah wants to. Point to where Hannah wants to!
Test 2	Full screen: Alli pouring milk.	Alli is pouring milk. Look, Alli is pouring milk! See, Alli is pouring milk.	Alli is pouring a drink. Look, Alli is pouring a drink! See, Alli is pouring a drink.
	Split screen: Hannah drinking milk; Hannah pouring juice.	Point to the picture where Hannah wants some! Show me the picture where Hannah wants some. Point to where Hannah wants some!	Point to the picture where Hannah wants to! Show me the picture where Hannah wants to. Point to where Hannah wants to!
Test 3	Full screen: Alli walking a toy dinosaur in front of her.	Alli is walking a dinosaur. Look, Alli is walking a dinosaur! See, Alli is walking a dinosaur.	Alli is walking a toy. Look, Alli is walking a toy! See, Alli is walking a toy.
	Split screen: Hannah kissing a dinosaur; Hannah walking bear.	Point to the picture where Hannah wants one! Show me the picture where Hannah wants one Point to where Hannah wants one!	Point to the picture where Hannah wants to! Show me the picture where Hannah wants to. Point to where Hannah wants to!
Test 4	Full screen: Alli bouncing stuffed cow on her lap.	Alli is bouncing a cow. Alli is bouncing a cow! Wow, Alli is bouncing a cow.	Alli is bouncing an animal. Alli is bouncing an animal! Wow, Alli is bouncing an animal.
	Split screen: Hannah bouncing stuffed lion; Hannah turning (in a circle, facing forward) a stuffed cow.	Point to the picture where Hannah wants one! Show me the picture where Hannah wants one Point to where Hannah wants one!	Point to the picture where Hannah wants to! Show me the picture where Hannah wants to. Point to where Hannah wants to!

Discourse Structure and Syntactic Parallelism in VP Ellipsis

Christina Kim and Jeffrey Runner

University of Rochester

1. Introduction

This paper presents psycholinguistic data from three magnitude estimation experiments that suggest a syntactic parallelism condition on VP ellipsis (VPE) that can be modulated by discourse coherence, and applies to VPE both within and across sentences. Section 1 will outline some of the arguments for and against syntactic and semantic accounts of VPE, and lay out the predictions of existing proposals with respect to syntactic mismatch in VPE. Section 2 reviews previous psycholinguistic studies on VPE. The three magnitude estimation experiments are presented in sections 3-5. Experiment 1 looks at cases of voice and category mismatch in VPE with sentential coordination, replicating some basic contrasts found by Arregui et al. (2006). Experiment 2 compares voice-mismatched VPE under different discourse coherence relations (Kehler 2000), and Experiment 3 compares voice mismatch in VPE in coordinate structures and across sentence boundaries.

Together, these findings support the existence of a syntactic parallelism condition on VPE which applies at the level of discourse structure. The observation that syntactic mismatch effects are observed to the same extent across discourse and within individual syntactic structures suggests that at least certain syntactic alternations, like passive, are encoded at the level of discourse representations. Empirical consequences of assuming different sources for mismatch-related unacceptability are discussed in section 6.

1.1 Licensing VP ellipsis: The identity condition

How are sentences with ellipsis interpreted? In sentences containing various types of ellipsis—like VPE (1a), sluicing (1b), pseudogapping (1c),

comparative deletion (1d)—it's clear that the understood meaning of the elided piece is similar to that of an overt phrase elsewhere in the sentence. We'll refer to this overt phrase as the antecedent.

- (1) a. Jill made fun of Abby, and Matt did ~~make fun of Abby~~, too.
 b. The department really wanted to hire a psycholinguist straight out of grad school, but I don't remember who ~~the department really wanted to hire straight out of grad school~~.
 c. Sameer picked out all the red jelly beans, and Justin did ~~pick out the purple ones~~.
 d. Jane smokes more cigarettes per day than Nate does ~~smoke x cigarettes per day~~.

In particular, the elided material can't be construed as something else; for example, (1d) cannot mean that Jane smokes more cigarettes per day than Nate drinks cups of coffee. The exact nature of the similarity that must obtain between antecedent and ellipsis site has been under debate.

1.2 Syntactic identity

Before asking what kind of identity must hold between antecedent and elided VPs, we might ask what kinds of information are in the ellipsis site. It's in principle possible that the identity requirement holds across just one of the types of information encoded in the reconstructed VP representation—for instance, semantic or property-level identity could hold of representations that also encode rich syntactic information. However the reverse is not true: syntactic identity could not hold across representations that encode only semantic information.

In fact, there is reason to think the ellipsis site contains unpronounced syntactic structure in sentences like those in (1). When an elided constituent would contain an island violation if its syntactic structure were reconstructed into the ellipsis site, a sentence becomes unacceptable (Haik 1987; Johnson 2001). This is illustrated for an adjunct island in (2b) (from Johnson, 2001) and a *wh*-island in (3b) (from Kennedy, 2003). If VPE involves deletion or non-pronunciation of a piece of syntactic structure, the unacceptability can be attributed to the island violations in (2c) and (3c).

- (2) a. Sterling criticized every decision that Lou did Δ .
 b. Sterling criticized every decision that Doug was upset because Lou did Δ .
 c. Sterling criticized every decision *wh_i* that Doug was upset because Lou did ~~criticize~~ *wh_i*.
- (3) a. I know which book Mag read, and which book you hadn't Δ .
 b. I know which book Mag read, and which book Mr. Yunioshi asked why you hadn't Δ .
 c. I know which book Mag read, and *which book_i* Mr. Yunioshi asked why you hadn't ~~read~~ *which book_i*.

Is it these syntactic representations that must meet the identity condition? Some of the earliest accounts propose that syntactic identity must hold either at surface structure or logical form for VPE to be well-formed (Sag 1976; Williams 1977; Sag & Hankamer 1984, also Hankamer 1979; Tancredi 1992; Wilder 1995). More recent work (e.g. Merchant 2008; Arregui et al. 2006; Kobele 2006; Kennedy & Merchant 2000; Kennedy 2003) has taken up versions of syntactic identity as well. A strict syntactic identity condition accounts for the unacceptability of (4b), where the antecedent and ellipsis site differ in voice, relative to its matched counterpart (4a).

- (4) a. Jill betrayed Abby, and Matt did ~~betray~~ Abby, too.
 b. Jill betrayed Abby, and Matt was ~~betrayed by Jill~~, too.

But, as noted by Kennedy & Merchant (2000); Arregui et al. (2006); Tanenhaus & Carlson (1990), and others, purely syntactic accounts appear to undergenerate, to the extent that certain sentences with syntactic mismatch are acceptable (5).

- (5) a. This information could have been released by Gorbachov, but he chose not to. (*Daniel Shorr, NPR, 10/17/92, from Hardt 1993, cited in Arregui et al. 2006*)
 b. In March, four fireworks manufacturers asked that the decision be reversed, and on Monday the ICC did. (*from Rosenthal 1988; cited in Dalrymple et al. 1991, Kehler 2002*)
 c. This problem was to have been looked into, but nobody did ~~look into~~ it. (*from Kehler 2002, cited in Arregui et al. 2006*)

Thus, while sensitivity to syntactic phenomena like binding and passivization suggest that there is syntactic structure in the ellipsis site and that there should be a condition on VPE that is stated in terms of syntactic structure, a strictly syntactic identity condition leaves some apparently acceptable cases of syntactic mismatch unexplained.

1.3 Semantic identity

An alternative approach to licensing VPE allows for such sentences by requiring semantic identity between the ellipsis site and its antecedent (Dalrymple et al. 1991; Hardt 1993)—on this view the elided VP is syntactically like a proform, in that it has no internal syntactic structure, and picks up its meaning from the context of utterance. For Dalrymple et al. (1991), ellipsis is resolved through higher-order unification, where identity is at the level of property-denoting expressions. Thus (6a-6b) is understood by finding an antecedent property (6c) and substituting it into the second conjunct (6d-6e).

- (6) a. Someone released this information, but Gorbachov didn't Δ .
 b. *release(someone, this information) but not $P(\text{Gorbachov})$.*
 c. $P = \lambda x. \text{release}(x, \text{this information})$.
 d. *release(someone, this information)*
 but not $\lambda x. \text{release}(x, \text{this information})(\text{Gorbachov})$.
 e. *release(someone, this information)*
 but not $\text{release}(\text{Gorbachov}, \text{this information})$.

The problem is that such accounts will overgenerate—(7) should be resolved in the same way as (6), but here the voice mismatch seems to degrade acceptability.

- (7) This information was released, but Gorbachov didn't Δ .

1.4 Hybrid approaches: Discourse-level alignment

How should the need for strict syntactic identity in some cases and the need for syntactic flexibility in others be resolved? One approach to reconciling this apparent conflict is to take the above observations to indicate that both semantic and syntactic parallelism are available as ways of recovering the elided VP. The problem then becomes figuring out what properties define

the cases where the syntactic and semantic recovery mechanisms are used. This is the approach taken by Kehler (1995, 2000, 2002), who proposes that whether parallelism is syntactic or semantic depends on how the meanings of the clauses containing the two VPs are related.

For Kehler, whether syntactic effects appear depends on the discourse relation between the clauses containing the antecedent and the elided phrase. In particular, an elided expression in a sentence that is part of a Cause-Effect relation only needs to match its antecedent in propositional content, and shouldn't require syntactic information. On the other hand, the elided material in a Resemblance relation relies on aligning its syntactic arguments with those of its antecedent, and therefore should show degradation when there is syntactic mismatch.

Discourse coherence-based accounts predict interactions between Mismatch and Discourse relation type. (4b) and (9) should be worse than their Matched counterparts (4a) and (8), since they are instances of Parallel conjuncts.

- (8) The report criticized Roy, but Kate didn't ~~criticize~~ Roy
- (9) a. The criticism of Roy was harsh, but Kate didn't ~~criticize~~ Roy
b. The report was critical of Roy, but Kate didn't ~~criticize~~ Roy

But (11) and (13) should not be worse than their Matched counterparts (10) and (12) because they are instances of the Cause-Effect. No asymmetry is predicted between Active-Passive and Passive-Active sequences.

- (10) a. Jill invited Abby, so Matt did ~~invite~~ Abby, too.
b. Abby was invited by Jill, so Matt was ~~invited by Jill~~, too.
- (11) a. Jill invited Abby, so Matt was ~~invited by Jill~~, too.
b. Abby was invited by Jill, so Matt did ~~invite~~ Abby, too.
- (12) The report criticized Roy, so Kate didn't ~~criticize~~ Roy
- (13) a. The criticism of Roy was harsh, so Kate didn't ~~criticize~~ Roy
b. The report was critical of Roy, so Kate didn't ~~criticize~~ Roy

Mixed approaches to ellipsis resolution such as Kehler's have the advantage that they allow for both syntactic and semantic recovery mechanisms, and specify the discourse conditions under which one mechanism is

used other another. Such a theory links observed unacceptability due to violations of syntactic or semantic identity to the use of a particular (syntactic or semantic) VP recovery mechanism, which is determined categorically by the type of discourse context the ellipsis is embedded in. Thus the strongest form of Kehler's proposal seems to predict sensitivity to violations of syntactic identity under the Resemblance relation, and complete insensitivity to syntactic structure under Cause-Effect.

2. Experimental evidence

Tanenhaus & Carlson (1990) provide some initial psycholinguistic evidence in favor of a syntactic condition on VPE. In a makes-sense judgment task, they compared how comprehenders' judgments were affected by syntactic mismatch in sentences containing VPE (*surface anaphora*; Hankamer & Sag 1976; Sag & Hankamer 1984) or 'do it' (*deep anaphora*). They discovered that in VPE, pairs of sentences where the antecedent and elided material did not match in voice (14b-14c) or category (15b-14c) were judged to make sense a smaller proportion of times than when there was syntactic match (14a-14c), (15b-14c). On the other hand, the proportion of sentence pairs judged to make sense did not differ between deep anaphora sentences (14d, 15d) where the two sentences either matched or mismatched in voice (14a-14b) or category (15a-15b).

- (14) a. Someone had to take out the trash. ...
 b. The trash had to be taken out. ...
 c. ... But Bill refused to \triangle .
 d. ... But Bill refused to do it.
- (15) a. It always annoys Sally when anyone mentions her sister's name....
 b. The mention of her sister's name always annoys Sally. ...
 c. ... However, Tom did \triangle anyway out of spite.
 d. ... However, Tom did it anyway out of spite.

It appears that in the case of 'do it' anaphora, semantic match between the antecedent and anaphoric VPs is sufficient to license the anaphoric dependency—the 'do it' VP can pick up its meaning from the antecedent whether or not it matches syntactically. In contrast, VPE seems to be sensitive to the syntactic match between the elided material and the antecedent VP.

While these results establish that VPE can show sensitivity to syntactic mismatch, they don't say anything about conditions under which sequences like (14b)-(14c) are acceptable. In fact, Kehler (2002); Hardt (1993); Dalrymple et al. (1991), and others have pointed out that there are naturally-occurring examples of syntactic mismatch in VPE (5). The antecedent VPs in (5) all mismatch their respective elided VPs in voice.

Perhaps more importantly, Tanenhaus & Carlson (1990) provide no account of the gradedness of acceptability, though it seems that acceptability degrades as the extent of the structural mismatch increases. Arregui et al. (2006)'s proposal tries to capture the relationship between acceptability and degree of syntactic mismatch. They asked participants to rate the acceptability of sentences like (16), which were assumed to vary by how syntactically similar the antecedent was to the elided material.

- (16) a. None of the astronomers saw the comet, but John did \triangle .
 b. Seeing the comet was nearly impossible, but John did \triangle .
 c. The comet was nearly impossible to see, but John did \triangle .
 d. The comet was nearly unseeable, but John did \triangle .

Because they assume a strict syntactic identity requirement on VPE, all the sentences in (16) are ungrammatical for Arregui et al. (2006) except (16a). To comprehend such sentences, they propose that the comprehender applies special repair operations to the mismatching antecedent, which eventually yield an antecedent that meets the syntactic identity requirement. This addresses the graded acceptability problem by allowing certain ungrammatical sentences to be parsable with the help of repair operations; the greater the number of repair steps required to restore syntactic well-formedness, the less acceptable a sentence sounds.

While both Tanenhaus & Carlson (1990) and Arregui et al. (2006) provide experimental evidence that VPE is sensitive to syntactic non-identity in the form of voice and category mismatches, neither compares the critical ellipsis sentences to their no ellipsis counterparts. As such, the possibility that the observed contrasts are due to a general dispreference for non-parallel conjuncts (instead of an ellipsis-specific parallelism condition) cannot be excluded.

3. Experiment 1: Voice and Category mismatch in VPE

The first experiment looked at cases of VPE where the antecedent was not a perfect structural match for the elided VP. In critical conditions, the antecedent differed in either voice (active vs. passive) or category (nominal or adjectival vs. verbal) from the elided VP in the second conjunct. In addition, each Ellipsis condition had a corresponding No Ellipsis condition, to determine whether any syntactic mismatch effects were general non-parallelism effects, or specific to VPE licensing. The objective of Experiment 1 was to establish an empirical base that any model of VPE would minimally have to account for.

Based on previous experimental results (e.g. Tanenhaus & Carlson 1990; Arregui et al. 2006) we expect structural mismatch to result in degraded acceptability. However there are certain syntactic accounts of ellipsis, like Merchant (2008), that predict voice mismatch will not be subject to syntactic identity in VPE due to the low ellipsis height (compared to TP-level ellipsis). Under such an account, any decrease in acceptability in syntactic mismatch conditions will be unexplained.

If the mismatch cases are degraded relative to their syntactic match counterparts, there is a question of whether the syntactically more complex form will be easier to have as an antecedent to the syntactically simpler form, or vice versa. Chung (2006) proposes a “morphological subset” principle that states that each morpheme included in the ellipsis site must have a corresponding morpheme in the antecedent (but not vice versa). According to this principle, Active-Passive would be predicted *worse* than Passive-Active sequences, assuming that the elided constituent in a Passive VP contains more morphological features than the corresponding Active VP. Arregui et al. (2006) in fact find that Active-Passive is rated worse than Passive-Active; they suggest a memory-based explanation involving mis-remembering a mismatching antecedent as a matching one. This mis-remembering is supposed to be easier when a Passive is being mis-remembered as an Active, than the other way around, presumably because the Active form is more basic.

Both Chung (2006) and Arregui et al. (2006) suggest a connection between the morphosyntactic complexity of the passive form and its relative unacceptability in the ellipsis site, though the explanations provided for this asymmetry are different. We know of no proposal that predicts

an asymmetry in the reverse direction (Active-Passive predicted better than Passive-Active order).

3.1 Design

There were 8 cells in the experiment (Ellipsis (Ellipsis, No Ellipsis) x Mismatch (Match, Mismatch) x Mismatch Type (Voice, Category)). Half of the Voice mismatch cases had Active (antecedent)-Passive (ellipsis) order, and half had Passive-Active order. Half of the Category mismatches had a Nominal antecedent-VPE sequence, and half had an Adjectival antecedent-VPE sequence. Example sentences for the experimental conditions are given in Table 1.

		Match	Mismatch
No ellipsis	Voice	Abby invited Bill, and Jill invited him, too	Abby invited Bill, and Jill was invited by her, too
	Category	Abby criticized Bill, and Jill criticized him, too	The criticism of Bill was harsh, but Jill didn't criticize him
Ellipsis	Voice	Abby invited Bill, and Jill did, too	Abby invited Bill, and Bill was, too
	Category	Abby criticized Bill, and Jill did, too	The criticism of Bill was harsh, but Jill didn't

Table 1: Experiment 1 design and example stimuli.

3.2 Method and Procedure

The experimental paradigm used was magnitude estimation, adapted from e.g. Bard et al. (1996). In this paradigm, participants give numerical ratings to stimuli relative to the rating they gave to some standard, or modulus, at the beginning of the experiment. For language stimuli, the ratings are participants' estimates of the acceptability of the sentence in the current trial compared to the acceptability of the modulus.

The current experiment was run on a Macintosh computer running PsyScope software. Participants first practiced giving estimates of line lengths (cf. Bard et al. 1996), then practiced with sentences. Then they assigned a value to the modulus sentence: *'The children were amused by the cartoon, but their parents weren't'*. On each trial, the modulus appeared on the screen

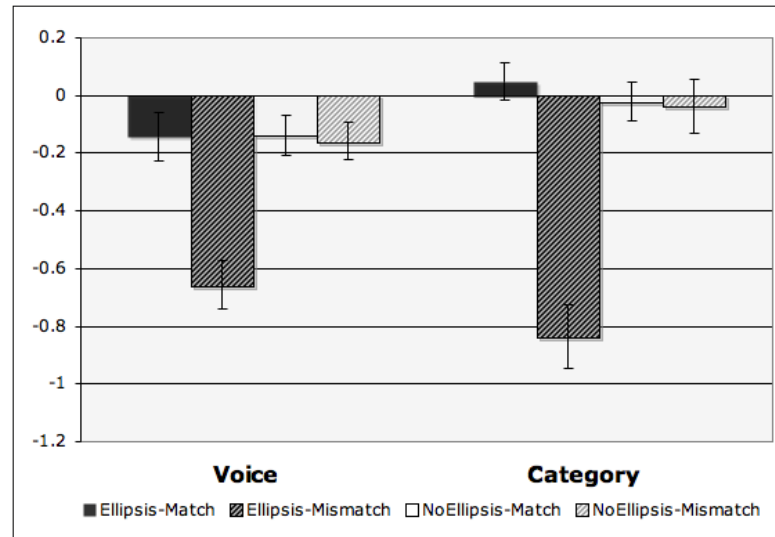


Figure 1: Experiment 1 results (error bars are Standard Error).

together with the sentence to be rated on that trial. Participants typed their estimates into a text box, then pressed the spacebar to proceed to the next trial.

There was one break halfway through the trials, and the whole experiment took participants approximately 10-15 minutes. 20 native English speakers from the University of Rochester community participated.

3.3 Results

3.3.1 Data analysis

The data were first normalized by dividing each participant's estimates by their modulus value. All analyses were performed on log-transformed values of the normalized data.

The log scores were fitted to a linear mixed-effects model, with three fixed effects (Ellipsis, Mismatch, Mismatch Type), and one random effect (Subject). This model was selected from among other linear models containing subsets of the experimental factors as fixed effects, by comparing the models' log likelihoods. The results of Experiment 1 are plotted in Figure 1.

3.3.2 Main effects and interactions

There were main effects of Ellipsis ($F(1, 19) = 51.0, p < .0001$), Mismatch ($F(1, 19) = 99.2, p < .0001$), and Mismatch Type ($F(1, 19) = 4.5, p < .05$). Ellipsis was worse than NoEllipsis ($t = 10.1, p < .0001$), and Mismatch was worse than Match ($t = 12.0, p < .0001$). In addition, Voice mismatches were worse than Category mismatches ($t = 1.9, p = .05$).

	Ellipsis	No Ellipsis	Match	Mismatch	Voice	Category
<i>mean</i>	-.4	-.09	-.42	-.06	-.27	-.21
<i>st.dev.</i>	.58	.33	.57	.32	.43	.56

Table 2: Experiment 1 main effects.

In addition, there was an Ellipsis-Mismatch interaction ($F(1, 19) = 61.7, p < .0001$), such that Mismatch was worse than Match only when there was Ellipsis (for Ellipsis conditions, Mismatch is worse than Match ($t = 16.4, p < .0001$); for NoEllipsis, Mismatch \approx Match ($t = .6, p = 1.0$)). Mean estimates for each factor are given in Table 2, and condition means in Table 3.

		No Ellipsis/ Match	No Ellipsis/ Mismatch	Ellipsis/ Match	Ellipsis/ Mismatch
Voice	<i>mean</i>	-.14	-.16	-.14	-.66
	<i>st.dev.</i>	.30	.29	.37	.47
Category	<i>mean</i>	-.02	-.04	-.05	-.84
	<i>st.dev.</i>	.29	.40	.28	.63

Table 3: Experiment 1, mean log estimates by condition.

3.3.3 Comparisons by Mismatch Type

The Ellipsis trials for Voice and Category conditions were also analyzed separately in post hoc comparisons.

Within the Voice mismatch trials, the antecedent was either an active or a passive VP, yielding four possible active-passive orderings: Active-Active, Passive-Passive, Active-Passive, and Passive-Active. Active-Active

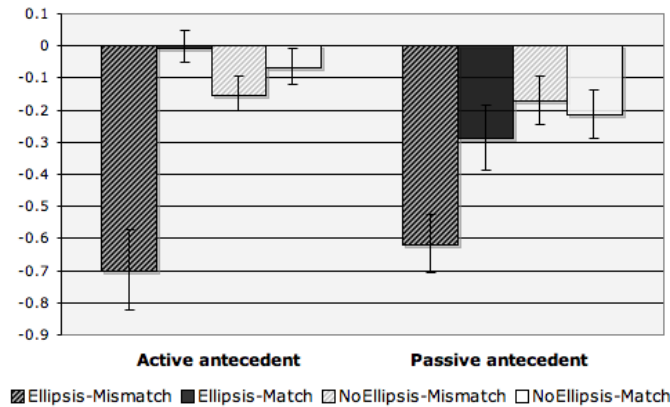


Figure 2: Experiment 1, Voice conditions. Mismatch and Match bars correspond to Active-Passive and Active-Active (left-hand side of graph), and Passive-Active and Passive-Passive (right-hand side).

sentences were rated better than Passive-Passive (corrected $p < .0001$), which were in turn better than both Active-Passive and Passive-Active (corrected $p < .0001$). The two mismatch conditions did not differ from each other (corrected $p = .29$). The results for Voice conditions only are in Figure 2.

For Category trials, the antecedent was an active VP in Match trials, and either a nominal or an adjectival phrase in Mismatch trials, yielding the following three orderings: VP-VP, NP-VP, and AP-VP. VP-VP sentences were better than NP-VP (corrected $p < .0001$), which were better than AP-VP (corrected $p < .05$). The results for Category conditions only are in Figure 3.

3.4 Discussion

Experiment 1 replicates the basic finding of Tanenhaus & Carlson (1990) and Arregui et al. (2006) that sentences with syntactic mismatch between the first and second conjuncts (or sentences) are judged less acceptable than sentences where conjuncts match syntactically. Importantly, this mismatch effect is only observed when there is ellipsis in the second conjunct; this suggests the sensitivity to syntactic parallelism reflects a constraint on VPE, not a general prohibition on syntactic mismatch in coordinate structures.

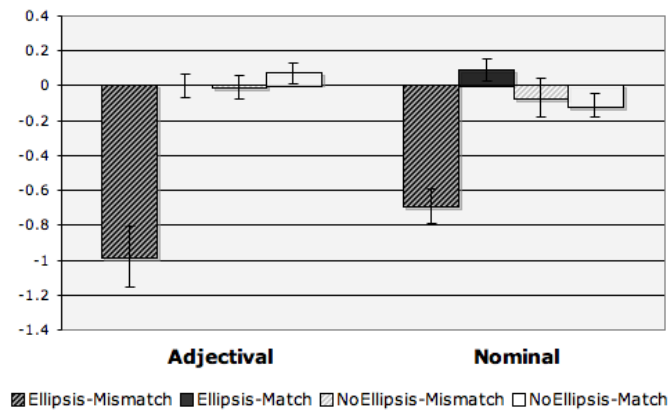


Figure 3: Experiment 1, Category conditions. Mismatch and Match bars correspond to AP-VP and VP-VP (left), and NP-VP and VP-VP (right).

To revisit what is predicted by some existing syntactic accounts of ellipsis in light of the results from Experiment 1, for Merchant (2007, 2008), asymmetries between VP-level ellipsis and TP-level ellipsis is captured by ellipsis height. In the case of voice mismatch, whether a particular kind of ellipsis will permit it depends on whether the ellipsis applies to a constituent in the syntactic tree that includes the Voice head. If the ellipsis is high, as in sluicing or gapping, and Voice is part of the elided material, it is subject to the syntactic identity condition, and voice mismatch will not be possible. If the ellipsis is low enough that Voice is external to the ellipsis site, as in VPE, the value of the passive feature on the Voice head escapes the syntactic identity condition and can differ between the antecedent and ellipsis conjuncts. While the results of Experiment 1 clearly show that syntactic mismatch results in decreased acceptability, it would be interesting to know whether the strength of the syntactic mismatch effect is in general related to the size of the ellipsis site.

According to both Chung's (2006) morphological subset principle and Arregui et al. (2006), Active-Passive order should have been rated less acceptable than Passive-Active order in sentences with ellipsis; more generally, the more morphologically or derivationally complex form was predicted to be worse in the ellipsis conjunct. While there was a numerical

difference that appeared to be in this direction for voice mismatches (Active-Passive order less acceptable than Passive-Active), this difference was not significant. However for the category mismatches, the Adjective-VP sentences were rated worse than the Nominal-VP sentences. We return to this issue in the general discussion.

A possibility not excluded by Experiment 1 is that syntactic mismatch effects associated with VPE are limited to cases where the conjuncts are in specific discourse relations (Kehler 2000, 2002). In particular, the sensitivity to syntactic parallelism may be restricted to instances of the Resemblance relation. If this is the case, changing the coherence relation to one that doesn't depend on syntactic parallelism, like Cause-Effect, should eliminate any effects of mismatch. Experiment 2 tests this prediction by manipulating discourse coherence (Resemblance vs. Cause-Effect) in addition to syntactic mismatch.

4. Experiment 2: Syntactic effects are modulated by Discourse Coherence

Experiment 2 compares VPE sentences with voice mismatch, when the discourse relation between the antecedent and the ellipsis conjuncts was either Resemblance or Cause-Effect. According to Kehler (2000, 2002), the Resemblance relation relies on the alignment of arguments from one sentence (or conjunct) to the next, and as such should be sensitive to changes in syntactic parallelism. On the other hand, the Cause-Effect relation relates sentences at the propositional meaning level, and should therefore be insensitive to syntactic manipulations that leave the meaning contribution of each conjunct intact.

Frazier & Clifton (2006) have a similar set of experiments where they compare sentences with ellipsis where the relation between the antecedent and ellipsis conjuncts is either Resemblance or Cause-Effect. In both percent 'got it' data and acceptability ratings on a 1 to 5 scale, they failed to find the asymmetry predicted by Kehler: the Resemblance cases were no worse than the Cause-Effect cases. While they did show for the acceptability ratings that corresponding sentences with syntactically matching conjuncts did not differ—that is, the lack of difference in the mismatching sentences was not due to a difference in the acceptability of the matching forms—they did not show this for the no ellipsis counterparts. In addition, while they

themselves (based on Arregui et al. (2006)) would expect mismatches with Active-Passive order to be more unacceptable than Passive-Active order, all the stimuli in Frazier & Clifton (2006) (Experiments 1 and 2) involved passive antecedents. The current experiment, in addition to testing Kehler's predictions about syntactic parallelism, includes both syntactic match and no ellipsis conditions, as well as both active and passive antecedents.

4.1 Design

There were 8 cells in the experiment (Ellipsis (Ellipsis, No Ellipsis) x Mismatch (Match, Mismatch) x Discourse Relation (Resemblance, Cause-Effect)). Half of the Match trials had active VPs, and half had passive VPs. Similarly, half of the Mismatch trials have Active-Passive order, and the other half had Passive-Active order. As illustrated in the example stimuli in Table 4, the coherence relation was indicated by connectives like *and* or *but* for Resemblance, and *so* or *because* for Cause-Effect.

		Match	Mismatch
No ellipsis	Resemblance	Abby invited Bill, and Jill invited him, too	Abby invited Bill, and Jill was invited by her, too
	Cause-Effect	Abby invited Bill, so Jill didn't invite him	Abby invited Bill, so Jill wasn't invited by her
Ellipsis	Resemblance	Abby invited Bill, and Jill did, too	Abby invited Bill, and Jill was, too
	Cause-Effect	Abby invited Bill, so Jill did, too	Abby invited Bill, so Jill wasn't

Table 4: Experiment 2 design and example stimuli.

4.2 Method and Procedure

The experimental paradigm and procedure were identical to Experiment 1, except that experimental trials were interspersed with filler sentences, which were either monoclausal or contained a discourse relation that did not appear in the test items.¹ 24 native English speakers from the Rochester

¹ Experiment 1, which included different types of category mismatches in addition to voice mismatches, had sufficient variation in the test items that we decided against including additional fillers.

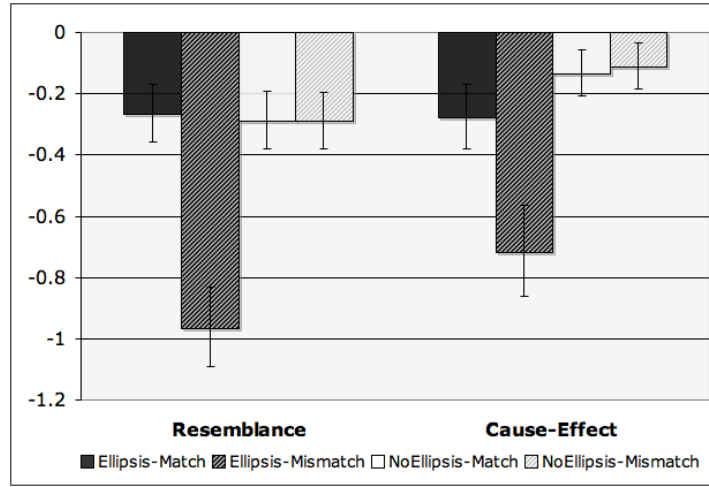


Figure 4: Experiment 2 results (error bars are Standard Error).

community participated in the experiment. Each experimental session took approximately 15 minutes.

4.3 Results

4.3.1 Data analysis

As in Experiment 1, estimates were normalized and log-transformed. The log scores were fitted to a linear mixed-effects model with three fixed effects (Ellipsis, Mismatch, Discourse Relation), and one random effect (Subject). The results of Experiment 2 are plotted in Figure 4.

4.3.2 Main effects and interactions

There were main effects of Ellipsis ($F(1, 23) = 119.92, p < .0001$), Mismatch ($F(1, 23) = 73.06, p < .0001$), and Discourse Relation ($F(1, 23) = 22.69, p < .0001$). As in Experiment 1, Ellipsis was worse than No Ellipsis ($t = 7.38, p < .0001$), and Mismatch was worse than Match ($t = 5.61, p < .0001$). In addition, Resemblance was worse than Cause-Effect ($t = 1.95, p = .05$). Mean log estimates are given by factor in Table 5.

Again as in Experiment 1, there was an Ellipsis-Mismatch interaction ($F(1, 23) = 102.91, p < .0001$), such that Mismatch was worse than

	Ellipsis	No Ellipsis	Match	Mismatch	Resemblance	Cause-Effect
<i>mean</i>	-.53	-.21	-.25	-.50	-.41	-.31
<i>st.dev.</i>	.33	.53	.35	.54	.46	.47

Table 5: Experiment 2 main effects.

Match for Ellipsis trials ($t = 8.58, p < .0001$), but not for No Ellipsis ($t = -.70, p = .49$). In addition, Mismatch interacted with Discourse relation ($F(1, 23) = 4.92, p < .03$)—the Mismatch effect was greater for Resemblance ($t = -5.4, p < .0001$) than for Cause-Effect ($t = -3.28, p = .001$). There was also a three-way Ellipsis-Mismatch-Discourse Relation interaction ($F(1, 23) = 3.95, p < 0.05$)—the Ellipsis-Mismatch interaction was stronger for Resemblance than for Cause-Effect. Condition means are given in Table 6.

		No Ellipsis/ Match	No Ellipsis/ Mismatch	Ellipsis/ Match	Ellipsis/ Mismatch
Resemblance	<i>mean</i>	-.29	-.29	-.26	-.96
	<i>st.dev.</i>	.36	.35	.49	.28
Cause-Effect	<i>mean</i>	-.13	-.11	-.28	-.71
	<i>st.dev.</i>	.28	.28	.40	.55

Table 6: Experiment 2, mean log estimates by condition.

4.3.3 Active-Passive order

One question brought up by these results is why Frazier & Clifton (2006) failed to find any coherence-related effects on acceptability, in a similar study manipulating voice mismatch in VPE. Recall that all of the syntactic mismatch stimuli in Experiments 1 and 2 of Frazier & Clifton (2006) have Passive-Active order. In fact, when we analyzed only the subset of Passive-Active sentences in our data, there was a numerical but not significant advantage of Resemblance-Ellipsis over Cause-Effect-Ellipsis ($t = .92, p = .36$). This numerical difference still contributes to the main effect of Discourse Relation Type, but it's evident that this effect is being carried

by the Active-Passive subset ($t = 2.32, p = .031$). Thus while we are still left with the interesting question of why reconstructing a passive from an active antecedent is more sensitive to the discourse context in which violations of syntactic identity occur, we may be able to reconcile the difference between the current findings and those reported in Frazier & Clifton (2006).

4.4 Discussion

Experiment 2 appears to be incompatible with a literal interpretation of Kehler (2000). But we interpret the interaction between discourse relation type and the size of the mismatch effect in the spirit of Kehler's proposal, by noting that the syntactic licensing condition on VPE seems to be more or less strictly enforced depending on the discourse context.²

The results of Experiments 1 and 2 together show that VPE is sensitive to syntactic parallelism, but that the extent of this sensitivity is modulated by the type of discourse coherence relation the antecedent and ellipsis conjuncts are part of. We note, however, that Kehler's theory of coherence types was intended to characterize relations between sentences, not two conjuncts in a coordinate structure. While it seems an intuitive extension of Kehler's original proposal to consider coherence relations between clauses within a sentence,³ there may be reasons not to do so.

A typical assumption in theoretical syntax is that core syntactic operations and principles are confined to the sentence domain—for instance, when considering possibilities for linguistic coreference, candidate referents outside the current sentence are not thought of as being in a c-command relationship with anaphors inside the sentence. Instead, discourse structural considerations are often assumed to play a more important role in relating sentences to each other in terms of their informational organization. Under this view, while the internal syntactic structure of one sentence may

² While we characterize the results of Experiment 2 in terms of Kehler's theory of discourse coherence (i.e. what we intended to manipulate), the current data do not allow us to distinguish this from Frazier & Clifton (2006)'s alternative hypothesis, which invokes a notion of parallelism distinct from the parallelism introduced by a Resemblance coherence relation. Specifically, they suggest that the presupposition introduced by the sentence-final '*too*' in many Resemblance sentences is a source of parallelism effects in VPE. Since information carried by the connective and '*too*' become available at different locations in the sentence, future experiments using online measures may help separate out the timecourses of these potentially different sources of parallelism.

³ See e.g. Rohde et al. (2008), who show that discourse coherence between clauses of a sentence can influence syntactic attachment preferences for a relative clause in the same sentence.

not affect the interpretation of subsequent sentences, it may contribute to the discourse structure by establishing what is given, focused, the topic or question under discussion. This in turn influences subsequent interpretation.

This view is taken up by Frazier & Clifton (2005), who claim on the basis of acceptability experiments that syntactic effects are limited in their domain to the sentence; once outside the sentence domain, a different set of interpretive constraints (based on extracting the main assertion from a sentence) applies at the level of discourse structure. They compared sentences like (17a), where an elided VP and its antecedent are in a single sentence with coordination, with pairs of sentences like (17b), where the antecedent VP is in one sentence and an elided VP is in another.

- (17) a. John said that Fred went to Europe and Mary did, too.
 b. John said that Fred went to Europe. Mary did, too.

Frazier & Clifton (2005) find that in the coordination condition, people are more likely to construe the ‘*go to Europe*’ VP as the antecedent of the elided VP, while in the two-sentence condition, they are less likely to consider it the antecedent.

If Frazier and Clifton are right that syntactic effects are confined to the sentence domain because comprehenders do not retain detailed syntactic representations across sentences, the syntactic mismatch effects observed in Experiments 1 and 2 are predicted to go away if the dependency between an elided VP and its antecedent crosses a sentence boundary. But following the basic intuition in Kehler’s work—that the relations between meanings in a discourse context influence whether structural or semantic information is important for interpreting ellipsis given an antecedent—we might expect such discourse-modulation to extend straightforwardly from clause sequences to sentence sequences. Experiment 3 tests this prediction.

5. Experiment 3: Syntactic effects above the ‘syntactic’ domain

Experiment 3 compares cases of VPE in a coordinate structure (Coordination) with VPE across a sentence boundary (Cross-sentential). Antecedent and ellipsis clauses were related by Resemblance or Cause-Effect, as in Experiment 2. The Coordination conditions are identical in structure to Experiment 1 (Voice conditions) and Experiment 2; as such we expect to see the same Ellipsis-Mismatch interaction pattern in the current experiment. The

critical question is whether there is an effect of syntactic mismatch when the antecedent of VPE is in the sentence preceding the sentence with ellipsis.

5.1 Design

There were 8 cells in the experiment (Ellipsis Type (Coordination, Cross-sentential) x Discourse Relation (Resemblance, Cause-Effect) x Mismatch (Match, Mismatch)). All of the Mismatch trials had voice mismatch; half of these had Active-Passive order, and half had Passive-Active order. All sentences contained ellipsis. The design of Experiment 3 and example stimuli are given in Table 7.

		Match	Mismatch
Coordination	Resemblance	Abby invited Bill, but Jill didn't.	Abby invited Bill, but Jill wasn't.
	Cause-Effect	Abby invited Bill, so Jill did, too.	Abby invited Bill, so Jill was, too.
Cross-sentential	Resemblance	Abby invited Bill. Jill did, too.	Abby invited Bill. Jill was, too.
	Cause-Effect	Abby invited Bill. So Jill did, too.	Abby invited Bill. So Jill was, too.

Table 7: Experiment 3 design and example stimuli.

5.2 Method and Procedure

The procedure and method were identical to Experiments 1 and 2. Approximately half of the practice and filler items contained two sentences. 14 native English speakers from the Rochester community participated. Each experimental session took approximately 15 minutes.

5.3 Results

5.3.1 Data analysis

As before, data were normalized and log-transformed. Log scores were fitted to a linear mixed-effects model with three fixed effects (Ellipsis Type, Discourse Relation, Mismatch) and one random effect (Subject).

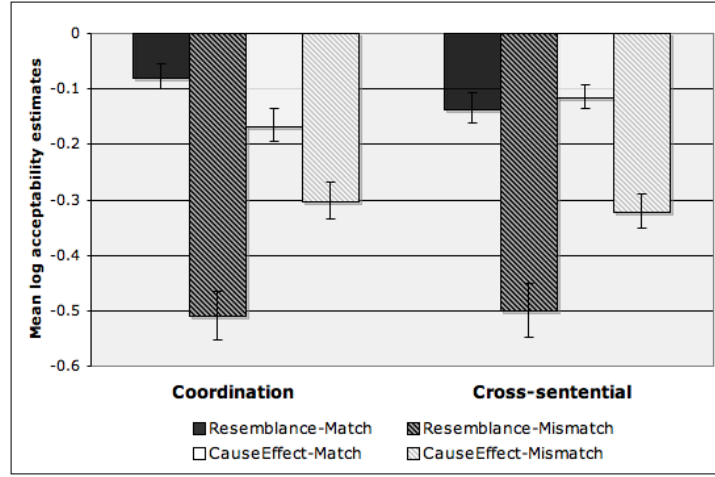


Figure 5: Experiment 3 results (error bars are Standard Error).

5.3.2 Main effects and interactions

As in the previous two experiments, there was a main effect of Mismatch ($F(1, 13) = 187.69, p < .0001$): Mismatch was worse than Match ($t = 11.87, p < .0001$). There was also a main effect of Discourse Relation ($F(1, 13) = 14.56, p = .0001$), as in Experiment 2: Resemblance was worse than Cause-Effect ($t = 3.08, p < .005$). There was no main effect of Ellipsis Type ($F(1, 13) = .034, p = .85$): VPE in Coordination did not differ from Cross-sentential VPE.

	Match	Mismatch	Resemblance	Cause-Effect	Coord.	Cross-sentence
<i>mean</i>	-.28	-.94	-.70	-.52	-.61	-.62
<i>st.dev.</i>	.53	.86	.90	.64	.79	.79

Table 8: Experiment 3 main effects.

There was a Discourse Relation-Mismatch interaction ($F(1, 13) = 30.25, p < .0001$), as in the previous experiment. VPE was more degraded under Resemblance than Cause-Effect, when there was Mismatch between the two clauses ($t = 4.89, p < .0001$), but not when there was Match ($t = 1.37, p = .17$). No two-way or three-way interactions involving Ellipsis

Type were significant. In other words, whether VPE occurred within a single sentence or across two sentences had no effect on acceptability. The data are plotted in Figure 5. Mean scores are given by factor in Table 8; condition means are in Table 9.

		Resemblance/ Match	Resemblance/ Mismatch	Cause-Effect/ Match	Cause-Effect/ Mismatch
Coordination	<i>mean</i>	-.18	-1.17	-.38	-.69
	<i>st.dev.</i>	.46	.92	.62	.68
Cross-sentential	<i>mean</i>	-.31	-1.15	-.26	-.74
	<i>st.dev.</i>	.55	1.03	.45	.65

Table 9: Experiment 3, mean log estimates by condition.

5.3.3 Active-Passive order

We again separately analyzed the subset of Mismatch trials by Active-Passive order; the Mismatch data were fit to a linear mixed-effects model with Active-Passive order and Discourse Relation as fixed effects, and Subject as a random effect. As in Experiment 2, there was an effect of order: Active-Passive order was worse than Passive-Active order when there was Mismatch ($F(1, 13) = 3.87, p < .01$). Interestingly, this order effect interacted with Discourse Relation ($F(1, 13) = 13.04, p < .0001$): Active-Passive order showed greater degradation under Resemblance relative to Cause-Effect (mean difference = .53), compared to Passive-Active order (mean difference = .36).

5.4 Discussion

Experiment 3 finds that the modulation of structural mismatch by discourse coherence when interpreting VPE (Experiment 2) extends to cases where the antecedent-ellipsis dependency is across sentences. This result argues against a syntax/discourse divide as proposed by Frazier & Clifton (2005): the Ellipsis-Mismatch interaction was unaffected by whether antecedent and ellipsis site were in the same sentence, or separated by a sentence boundary.

The fact that we find syntactic mismatch degrades VPE even across sentences seems to indicate that we have access to syntactic information

at the level of discourse: as a sentence containing VPE is interpreted, the antecedent-ellipsis relation is constrained by syntactic parallelism irrespective of whether the antecedent is part of the same syntactic structure or a different structure. If sentence interpretation were limited in such a way that only one unit of syntactic structure could be attended to at a time, there would be no way to compare the structure of an antecedent in a previous sentence to structure in the current sentence. In fact, it appears that any view of the relationship between syntactic and discourse structures where discourse representations contain no or very impoverished syntactic information will be unable to explain how syntactic identity can be enforced across discourse.

6. General discussion

6.1 Reconstructing elided VPs under syntactic identity

The results from Experiments 1-3 (and previous work by Tanenhaus & Carlson (1990) and Arregui et al. (2006), and work in the syntax literature (e.g. Kennedy 2003; Johnson 2001; Merchant 2007, 2008) show that there has to be a syntactic licensing condition on VPE. But even if we accept that there is a syntactic identity condition on VPE, there is the further question of what we assume happens when there fails to be perfect syntactic match between the antecedent and elided VPs. Since we observe that even in the face of syntactic mismatch, comprehenders are able to understand what a sentence means, there must be a way to parse and interpret such a sentence that involves a mechanism for restoring or preserving syntactic identity.

Arregui et al. (2006)'s proposal tries to capture the relationship between acceptability and degree of syntactic mismatch. Because they assume a strict syntactic identity requirement on VPE, all the sentences in (16) are ungrammatical for Arregui et al. (2006) except (16a). To comprehend such sentences, they propose that the comprehender applies special repair operations to the mismatching antecedent, which eventually yield an antecedent that meets the syntactic identity requirement. This addresses the graded acceptability problem by allowing certain ungrammatical sentences to be parsable with the help of repair operations; the greater the number of repair steps required to restore syntactic well-formedness, the less acceptable a sentence sounds.

In contrast, Kim et al. (2011) and Kobele et al. (2008) assume that

cases of syntactic mismatch are grammatical in that they are generated by the grammar. Instead, they link degradations in acceptability to the size of the elided constituent in the derivation tree. Syntactic mismatch will generally have the consequence that a smaller subpart of the antecedent and elided VP structures will be identical; this, combined with a parsing preference in the spirit of MaxElide (cf. Merchant 2008; Takahashi & Fox 2005) that prefers larger constituents to be elided, predicts the graded pattern of acceptability observed, with greater degrees of mismatch yielding greater degradation. A key difference from the approach in Arregui et al. (2006) is that syntactically non-parallel structures can be generated in a normal way, and as such do not have to invoke anything outside of the usual mechanisms for computing meanings from syntactic structures.

The empirical predictions of both repair and the ellipsis size accounts will ultimately depend on the grammar assumed to generate the structures in question. (Note that Arregui et al. (2006) claim all repair operations must correspond to existing syntactic operations specified by the grammar.) But in light of the findings to date, we seem to have at least one argument in favor of a repair-based account, and one in favor of a non-repair account.

One reason to prefer a repair account of VPE is that it provides a way to make sense of the asymmetry between Active-Passive and Passive-Active mismatches. Arregui et al. (2006); Kobele et al. (2008), and the results we report here all show that sentences with Active-Passive order are more degraded due to syntactic mismatch than their Passive-Active counterparts. Indeed, Chung (2006) and the Recycling Hypothesis proposed in Arregui et al. (2006) predict this asymmetry, since Active-Passive sentences require reconstructing a matching Passive structure from an Active antecedent. This draws on the observation that active syntax is more *basic* than passive syntax; under some analyses, a passive structure contains its corresponding active structure. While this is also true of the grammar assumed in Kobele et al. (2008), the absence of a repair process leaves no source for the observed mismatch asymmetry.

On the other hand, there appear to be no easy adjustments to repair accounts that would enable them to account for the discourse coherence effects we report here and in Kobele et al. (2008). From the point of view of Kobele et al. (2008), the discourse context that an instance of VPE appears in influences the strength of the ellipsis size constraint; since sentences con-

taining syntactic mismatch in VPE are generated by the grammar, there is no relationship between the extent to which this constraint is satisfied and the interpretability of the sentence. It is less obvious why changing the discourse context would affect the number of repair steps needed to restore syntactic identity (Arregui et al. 2006), and since syntactic identity is assumed to be strict, presumably VPE can only be interpreted after the full sequence of repair operations has been completed.

6.2 The relationship between syntax and the discourse context

A possible alternative explanation of the mismatch effects reported here that does not involve a syntactic identity condition is that what we are calling syntactic mismatch effects are really discourse structural mismatch. Since changes to syntactic structure (active vs. passive, nominal or adjectival vs. verbal) are likely to have corresponding effects on discourses (by means of changes in information structure, or predication structure), any of the structural mismatch effects reported here can be interpreted as discourse effects which are only indirectly related to syntactic structure. In fact, such an alternative is considered by Tanenhaus & Carlson (1990) from the point of view of comparing deep and surface anaphora.

It is difficult to pull apart effects of syntactic and discourse structure when they co-occur as they do here, but we note that whatever level of structure is responsible for the sensitivity to voice (or category) mismatches in VPE, it must be at least structurally rich enough that actives and passives (or nominalizations, deverbal adjectives, and verb phrases) have distinct representations. While the discourse might encode structure in other ways additional to the kinds of structural representations that distinguish actives from passives, we consider it plausible that the mismatch effects in Experiment 3 are syntactic effects, meaning that the discourse structure must also include syntactic structural representations. What we'll need in future work is a model of the discourse context that allows us to represent coherence relations, focus structure, and enough syntactic structure to capture the kinds of effects reported in this paper.

6.3 Interpreting VPE in discourses

Other issues that are outside the scope of this paper come up particularly when considering ellipsis across sentences. One is that there may be a lo-

cality bias for antecedents of ellipsis in situations where there is more than one candidate antecedent. An example of this ‘local antecedent’ bias might be (18). But note that the preference for the local antecedent (‘*apply for a job in Taipei*’) seems to be reduced in (19), which differs minimally from (18) in that the matrix verb has a more clearly eventive interpretation.

(18) I considered applying for a job in Taipei after Sameer did Δ .

(19) I decided to apply for a job in Taipei after Sameer did Δ .

While the factors that influence the choice of antecedent in multiple antecedent situations should be investigated further, it’s easy to imagine how such a bias might interact with syntactic mismatch effects. To take an example, consider (20a); to the extent that it’s easier to interpret it with the local antecedent, the local bias seems to reverse in (20b), where the local antecedent is a nominalization, but the non-local one is a normal VP.

(20) a. I thought about complaining to the registrar’s office after Sameer did Δ .

b. I thought about a complaint to the registrar’s office after Sameer did Δ .

Another discourse-level consideration that affects the strength of structural parallelism constraints is information structure: Kertz (2008) shows that, in addition to discourse coherence relations modulating lower level syntactic effects, the focus or information structure of a sequence of sentences may affect the relative ease of resolving ellipsis. Unlike the case of coherence relations, information structural effects found by Kertz appear to operate independently of ellipsis—in other words, there is a general preference for parallelism of coordinate structures or pairs of sentences with respect to focus structure. We leave to future research the interesting question of what the relationship is between various parallelism effects that result from general discourse pressures as opposed to narrower constraints specific to particular constructions or linguistic contexts.

References

- Arregui, Ana, Charles Clifton, Lyn Frazier & Keir Moulton (2006). Processing elided verb phrases with flawed antecedents: the recycling

- hypothesis. *Journal of Memory and Language* 55, pp. 232–246.
- Bard, Ellen Gurman, Dan Robertson & Antonella Sorace (1996). Magnitude estimation of linguistic acceptability. *Language* 72:1, pp. 32–68.
- Chung, Sandra (2006). Sluicing and the lexicon: The point of no return. Cover, R. T. & Y. Kim (eds.), *Proceedings of BLS 31: General Session and Parasession on Prosodic Variation and Change*, Berkeley Linguistics Society, Berkeley, CA, pp. 73–91.
- Dalrymple, Mary, Stuart M. Shieber & Fernando C.N. Pereira (1991). Ellipsis and higher-order unification. *Linguistics and Philosophy* 14, pp. 399–452.
- Frazier, Lyn & Charles Clifton (2005). The syntax-discourse divide: processing ellipsis. *Syntax* 8:2, pp. 121–174.
- Frazier, Lyn & Charles Clifton (2006). Ellipsis and discourse coherence. *Linguistics and Philosophy* 29, pp. 319–346.
- Haïk, Isabel (1987). Bound VPs that need to be. *Linguistics and Philosophy* 10, pp. 503–530.
- Hankamer, Jorge (1979). *Deletion in Coordinate Structures*. Garland, New York.
- Hankamer, Jorge & Ivan Sag (1976). Deep and surface anaphora. *Linguistic Inquiry* 7, pp. 391–428.
- Hardt, Daniel (1993). *Verb phrase ellipsis: Form, meaning, and processing*. Ph.D. thesis, University of Pennsylvania.
- Johnson, Kyle (2001). What VP ellipsis can do, and what it can't, but not why. Baltin, M. & C. Collins (eds.), *The Handbook of Contemporary Syntactic Theory*, Blackwell, pp. 439–479.
- Kehler, Andrew (1995). *Interpreting cohesive forms in the context of discourse inference*. Ph.D. thesis, Harvard University.
- Kehler, Andrew (2000). Coherence and the resolution of ellipsis. *Linguistics and Philosophy* 23:6, pp. 533–575.
- Kehler, Andrew (2002). *Coherence, reference and the theory of grammar*. CSLI Publications.
- Kennedy, Chris (2003). Ellipsis and syntactic representation. *The Syntax-Semantics Interface: Interpreting (Omitted) Structure*, John Benjamins.
- Kennedy, Chris & Jason Merchant (2000). Attributive comparative deletion.

- Natural Language and Linguistic Theory* 18:1, pp. 89–146.
- Kertz, Laura (2008). Verb phrase ellipsis in context. *Annual Meeting of the LSA*, Chicago.
- Kim, Christina, Gregory Kobele, Jeffrey Runner & John Hale (2011). The acceptability cline in verb phrase ellipsis. *Syntax* pp. xx–xx.
- Kobele, Gregory, Christina Kim, John Hale & Jeffrey Runner (2008). A processing model of ungrammatical VP ellipsis. Talk presented at 21st CUNY Conference on Human Sentence Processing, Chapel Hill.
- Kobele, Gregory M. (2006). *Generating Copies: An investigation into Structural Identity in Language and Grammar*. Ph.D. thesis, UCLA.
- Merchant, Jason (2007). Voice and ellipsis. Ms, University of Chicago.
- Merchant, Jason (2008). An asymmetry in voice mismatches in VP-ellipsis and pseudogapping. *Linguistic Inquiry* 39:1, pp. 169–179.
- Rohde, Hannah, Roger Levy & Andrew Kehler (2008). Implicit causality biases influence relative clause attachment. Poster presented at 21st CUNY Conference on Human Sentence Processing, Chapel Hill.
- Sag, Ivan (1976). *Deletion and Logical Form*. Ph.D. thesis, MIT.
- Sag, Ivan & Jorge Hankamer (1984). Toward a theory of anaphoric processing. *Linguistics and Philosophy* 7, pp. 325–345.
- Takahashi, Shoichi & Danny Fox (2005). MaxElide and the Re-binding Problem. Georgala, E. & J. Howell (eds.), *Proceedings of SALT 15*, CLC Publications, Ithaca, NY, pp. 223–240.
- Tancredi, Christopher (1992). *Deletion, deaccenting, and presupposition*. Ph.D. thesis, MIT.
- Tanenhaus, Michael & Greg Carlson (1990). Comprehension of deep and surface verb phrase anaphors. *Language and Cognitive Processes* 5, pp. 257–280.
- Wilder, Chris (1995). Some properties of ellipsis in coordination. *Geneva Generative Papers* 2, pp. 23–68.
- Williams, Edwin (1977). Discourse and logical form. *Linguistic Inquiry* 8, pp. 101–139.

Processing Null-Operator Structures*

Jason Overfelt

University of Massachusetts Amherst

1. Introduction

The resolution of a given syntactic chains could be achieved by one of two strategies. The parser could actively make predictions about the incoming material in an attempt to predict the positions of the chain links. Alternatively, the parser could employ bottom-up information in order to resolve the dependency using the material that has already been parsed.¹ Much of the literature on long-distance dependency resolution has converged on the view that the parser will immediately and predictively attempt to resolve a movement chain once the head of that chain has been identified. Furthermore, this task is given priority over other available parsing options. The idea that such a chain resolution strategy exists has been called the ACTIVE FILLER HYPOTHESIS (AFH; Frazier 1987; Frazier & Clifton Jr. 1989).²

Research on this topic has primarily focused on the resolution of WH-questions in English (Crain & Fodor 1985; Stowe 1986; Frazier & Clifton Jr.

*Thank you to all of those who participated in the experiments involved in this research. Thank you to thank the following people for their helpful comments about the work presented here: Rajesh Bhatt, Charles Clifton Jr., Brian Dillon, Lyn Frazier, Kyle Johnson, and Peggy Speas. And thank you to the participants of the Second-Year Seminar and to those who attended the Second-Year Mini-Conference, both at UMass. Any errors or misrepresentations of others' comments or ideas are solely my responsibility.

¹ See Fodor (1978) for a more fine-grained distinction of potential parsing strategies that will not be relevant for the purposes of this paper.

² There are several different theories concerning the impetus for an active search strategy as opposed to an arguably less error-prone strategy that employs only bottom-up information (Frazier & Clifton Jr. 1989; Pickering & Barry 1991; Pickering 1994; Gibson 1998; Aoshima et al. 2004). This paper will not explicitly argue for or against any one of the theories just mentioned. Instead, we will take note that the facts discussed in this paper suggest that the initiation of an active chain resolution strategy seems to be independent of particular lexical items or constructions. I will further address this point below.

1989; i.a.). Similar conclusion have also been reached for the processing of WH-questions in Dutch (Frazier 1987; Frazier & Flores D'Arcais 1989), German (Bader & Meng 1999), and Italian (De Vincenzi 1991) among others. More recent research has found that predictive parsing is also employed to resolve other leftward displacement operations such as Topicalization structures in Japanese (Miyamoto & Takahashi 2000) and English (Overfelt 2011) and Clitic Left-Dislocation (CLLD) structures in Brazilian Portuguese (Maia 1997) and Spanish (Pablos 2006). Interestingly, in these latter constructions the search is sensitive not to a gap, as in English WH-questions, but to the preverbal clitic.

However, there exist in natural languages unbounded, long-distance dependencies between syntactic positions that carry no phonological content. Such dependencies are typical in relative clause constructions (1) and are commonly thought to be headed by a null-operator that undergoes syntactic A'-movement to bind a variable in the extraction site (Chomsky 1980; Browning 1987).

- (1) $[_{DP} \text{The } [_{NP} [_N \text{letter}] \text{ } [_{CP} Op_1 (\text{that}) \text{ Sam wrote } x_1]]]$ is long.

It isn't obvious how the parser should or could parse an NOS. Here the parser is faced with the task of determining the scope of the operator and interpreting it with respect to the restricting Head without any direct evidence signaling the presence of the chain or the elements it is composed of.³ Without any explicit evidence for the chain, it seems likely that the parser would often not realize that the structure being interpreted actually contains a syntactic chain.

To my knowledge, the parsing of NOSs has received very little attention as far as processing is concerned.⁴ Alphonse & Davis (1997) predict that the parser, regardless of the environment in which the NOS appears, will not actively try to resolve the dependency chain. The idea is that the parser will never pursue an analysis that contains an element for which there is no phonetic evidence. The assumption I will take issue with is that the parser is concerned with the representational properties of the chain.

³ Even assuming a raising analysis (Vergnaud 1974; Kayne 1994) the point remains that there is no explicit evidence for the parser that the structure currently being parsed contains a movement chain.

⁴ There has however been some recent work done on WH-in-situ constructions in Mandarin (Xiang et al. 2010) and English (Sprouse et al. 2011) suggesting that the parser engages in a backward search to establish a proper scope taking position for an in-situ WH-element.

Experimentally, there has been done on post-verbal, WH-movement induced parasitic gap (PG) constructions in English.⁵ Various theoretical analyses have argued that the PG domain contains an NOS (Kayne 1983; Chomsky 1986; Browning 1987; Nissenbaum 2000).

(2) Which article did you read before filing?

Which article₁ did you read x_1 [*Op*₂ before filing *pg*₂]?

Early research focused just on the proposed accessibility hierarchy for the occurrence of parasitic gaps in Engdahl (1983: 9, ex. (7)).⁶ More recently however, Wagers & Phillips (2009) used moving window self-paced reading times to argue that the parser's default analysis of an adverbial clause did not include a parasitic gap despite its being licensed by a WH-element in the matrix clause. This study, while not explicitly interested in the parsing of NOSs, corroborates the predictions of Alphonse & Davis (1997).

I will show in what follows that NOSs are not always parsed by either an active or non-active resolution strategy. We will see that the issue of which strategy is employed is simply a matter of if and when in the parsing procedure the grammar requires the postulation of a chain. In this way, the resolution of dependency chains by the human sentence-processing mechanism is not driven by the links in the chain (i.e., the filler or the gap) but by the principles of the grammar itself. An active and predictive resolution strategy is employed only if the grammar requires a chain and this is apparent to the parser before encountering the final link in the chain. A non-active, bottom-up resolution strategy is employed when the grammar signals the parser that a chain is necessary and this is at or after the final link in the chain.

In a situation where the grammar provides an alternative parse not requiring an NOS, we will see that the parser first chooses to pursue the analysis that does not require the postulation of a chain. This idea could

⁵ We might also take into consideration the work on Italian null-subjects by De Vincenzi (1991).

⁶ This hierarchy is a proposed ranking of domains in which a PG was predicted to be acceptable. The primary division on the hierarchy followed the observation from (Ross 1967) that extraction out of a non-finite domain is in general more acceptable than extraction out of a finite domain. Early experimentation on PGs following (Engdahl 1983) was concerned with acceptability. At the time, it was still not widely accepted that PGs were grammatical constructions. Seely (1987) experimentally recreated this accessibility hierarchy for parasitic gaps through an acceptability task. Kurtzman & Crawford (1991) argued based on global reading times that parasitic gaps are more easily accessible when they are higher in the same hierarchy.

be conceptualized as the parser's attempt to mediate the demands of the grammar and the MINIMAL CHAIN PRINCIPLE (De Vincenzi 1991: 13).

(3) MINIMAL CHAIN PRINCIPLE (MCP)

Avoid postulating unnecessary chain members at S-structure, but do not delay required chain members.

As De Vincenzi claims, the MCP is applied in accordance with the grammar of a given language.⁷ This principle will guide the parser to first follow an available analysis that does not require an NOS. When the parser finds itself at a point in the parsing procedure where the only available parse provided by the principles of the grammar is one that requires an NOS, then the NOS will be postulated and resolved.

The remainder of this paper will focus on two specific goals. First, I will provide evidence for the claim that NOSs are resolved by way of both active and non-active parsing strategies depending on when in the parsing procedure the grammar makes the necessity of a chain apparent. This is the observation that leads us to suppose that the grammar directs the resolution of syntactic chains as opposed to the nature of the chain itself. Therefore, I will also be attempting to show that the application of one strategy or the other is a result of the constraints placed on the parser by the grammar.

To accomplish these goals, we will consider the potential four-way ambiguity that can arise with bare infinitival adjunct clauses in English.

(4) Sam took the towels [to wash in the river].

For now, we can note the primary ambiguity between having and not having an NOS. In the example above, this is the difference between interpreting wash in its intransitive frame or in its transitive frame and taking a variable as its argument. In section 2 we will see an explicit syntax for these readings.

Sections 3 and 4 present two experiments that test specific predictions about the preferred first-pass parse of these adjunct clauses and NOSs. Section 5 will interpret the results of these experiments in more detail and provide a discussion of the implications of this research. The first experiment in Section 3 is a self-paced reading study aimed at determining the initial

⁷ See De Vincenzi (1991, Ch. 7) for a discussion concerning the psychological motivation for a principle such as the MCP. She motivates this principle in terms of the most economic use of limited short term memory resources.

attachment preference for the adjunct clause. Following work on PP attachment ambiguities by Spivey-Knowlton & Sedivy (1995), I suggest that the parser will initially associate the adjunct clause as a modifier of the nominal or verbal domain as a function of the definiteness of the potential Head for the binding chain.⁸ Attachment into the nominal domain will commit the parser to the active and predictive construction with an NOS (a relative clause) whereas attachment into the verbal-domain does not.

The second experiment, which is also a self-paced reading study, is presented in Section 4. This experiment investigates further the behavior of the parser while interpreting an adjunct clause specifically when following a definite argument. Based on work in Pickering & Traxler (2003) we will interpret the presence of effects from the subcategorization preference of the verb in this environment to suggest that parser is not actively resolving a long-distance dependency. We will also see that the parser will overlook the availability of an NOS in the absence of bottom-up grammatical evidence requiring its presence.

2. The syntax of Bare Infinitival Adjunct Clauses

As the examples in (5) – (8) show, Bare Infinitival Adjunct Clauses (BIACs) in English are potentially four-ways ambiguous. The bracketed material is the bare infinitival adjunct clause. The paraphrases are meant to draw out the sometimes very similar interpretations that result from the possible structural configurations.

- (5) *Infinitival Relative Clause (IRC)*
 Sam₁ brought [DP a [NP [NP friend₂] [IRC λ_{x2} PRO₁ to show off x₂]]]
 “There is a friend that Sam could show off and Sam brought him.”
- (6) *Subject-gap Purpose Clause (SPC)*
 Sam₁ [VP brought [FP [DP a friend]₂ [VP] [SPC PRO₂ to show off]]]
 “There is a friend that Sam brought so the friend could show off.”
- (7) *Object-gap Purpose Clause (OPC)*
 Sam₁ [VP brought [FP [DP a friend]₂ [VP] [OPC λ_{x2} PRO₁ to show off x₂]]]

⁸ We will discuss this point in more detail in section 3, but for now we might assume that the parser loosely associates the adjunct to the root clause (cf. Frazier & Clifton Jr. 1989) and evaluates the available points of attachment with respect to the current discourse representation (cf. Crain & Steedman 1985; Altmann & Steedman 1988).

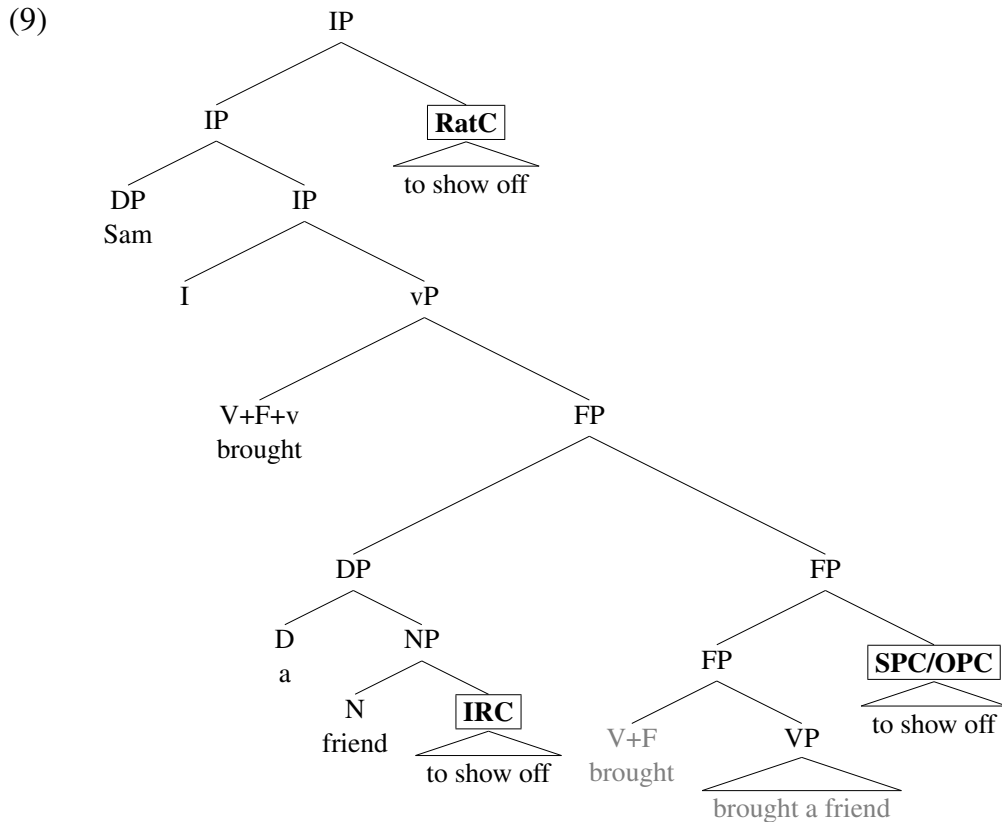
“There is a friend that Sam brought so Sam could show the friend off.”

(8) *Rationale Clause (RatC)*

[_{IP} [_{IP} Sam₁ brought a friend₂] [_{RatC} PRO₁ to show off]]

“There is a friend that Sam brought so that Sam could show off.”

The overall syntactic picture we will be assuming for these interpretations of the BIAC is shown below.



It's worth noting here that while the syntactic and semantic nature of the Head may differ from case to case, superficially there is very little about the internal syntax and semantics of the various BIACs that distinguishes them from each other.⁹ I assume with Huettner (1989) that this superficial

⁹ Various authors including von Stechow & Iatridou (2005), Bhatt (2006), as well as Hackl & Nissenbaum (accepted), have noted that each of these constructions expresses modal force to one degree or another. Additionally, all of these constructions involve a null-subject (PRO) and two of them (5) and (7) have null-operator structures. The differences we perceive between the construc-

similarity between the constructions above can be taken to indicate a deeper underlying similarity.

I will take the nominal-modifying IRC reading to result when the BIAC contains a null-operator chain and is merged into the nominal domain (5). I am assuming that a head-external complementation analysis is the correct syntax regardless of any further specifics concerning the elements involved in the IRC.

The literature from the era of Transformational Grammar (Faraci 1974), as well as the GB-era literature (Huettner 1989; Jones 1991) and more recent work (Whelpton 2002; Overfelt 2010) basically converge on an analysis of rationale clauses (8) as being adjoined to the matrix sentence or into the IP-domain. I take no real issue with the consensus and thus will simply adopt it into the analysis supposed here.

The same literature that has discussed RatCs has recognized their similarities to the purpose clauses in (6) and (7). Typically, the adjunct is assumed to be adjoined to the equivalent of the vP-domain. I will take an SPC and OPC to be adjoined to some functional projection (FP) between vP and VP (Overfelt 2010). The occurrence of a local A-movement (e.g. Johnson 1991) licenses the null-operator chain in OPCs or provides a binder for PRO in SPCs.¹⁰ In this way, the null-operator structure characteristic of the OPC is essentially a parasitic gap (in the sense of Nissenbaum (2000)) that is licensed by A-movement.¹¹

tions basically result from the same primary source, namely their point of attachment in the root clause. This in turn determines how the variables inside the BIAC will be interpreted and how they are interpreted with respect to the event variable(s) and modal operator(s) in the root clause (Overfelt 2010).

¹⁰ Another alternative analysis of OPCs does exist in the work of Nissenbaum (2005) where they are analyzed as event predicates that are directly selected by a sub-class of verbs. This argument is based on the observation that only predicates with rich event structures (e.g., a result state) are compatible with OPCs. See (Overfelt 2010) for evidence against such a position. Here we might simply note that the same facts are subsumed under an analysis claiming that object shift is permitted only given a rich event structure. Additionally, under a V-complement type of analysis we wouldn't predict the ordering facts shown above in (10) between the VP-attached locative adjunct from the kiosk and the OPC and we would lose the unifying properties of equating OPCs to SPCs.

¹¹ The consequences of this claim are neither few nor insignificant. Further discussion can be found in (Overfelt 2010), which includes a discussion of passivization, unergative and unaccusative predicates, as well as ECM and raising constructions.

3. Experiment 1

This section presents a self-paced reading study designed to determine the initial attachment preferences for the BIAC. Considering the ambiguity we saw in (5) – (8), the first step relevant to the parsing procedure will come after the parser has identified the string following the internal argument *the towels* to be a BIAC as opposed to a PP headed by *to*. For the purposes of this paper, we will assume that at this point in the parsing procedure, the BIAC is loosely associated to the root clause without any commitment to the domain of modification.¹² This treatment of the adjunct clause is comparable to the CONSTRUAL HYPOTHESIS proposed in (Frazier & Clifton Jr. 1996: 41).

(10) a. *Construal Principle*

- i. Associate a phrase XP that cannot be analyzed as instantiating a primary relation into the current processing domain.
- ii. Interpret XP within that domain using structural and non-structural (interpretive) principles.

b. *Current thematic processing domain*

The current thematic processing domain is the extended maximal projection of the last theta assigner.

The BIAC will be seen as a non-primary relation given that it is not a complement or subcategorized argument of the subject or the predicate. Given that the BIAC is not selected for by any constituent in the root clause, then we hypothesize that there is no generally preferred structural analysis. The parser will be left to determine the preferred domain of modification through interpretive principles.

In a study examining the parser's attachment preferences of PPs in cases of an ambiguous NP and VP modification, (Spivey-Knowlton & Sedivy 1995) found that PPs were preferably incorporated as NP modifiers when the DP to be modified was indefinite. When the same DP was definite,

¹² We will see the motivation later for the claim that the parser doesn't immediately commit to an attachment domain for the BIAC. For instance, it could be the case that the parsing principle LATE CLOSURE (Frazier 1978) pressures the parser to attach the BIAC into the phrase currently being processed: The DP *the towels*. Assuming that this is the case however forces us to make incorrect predictions concerning the definiteness effect alluded to in Section 1.

then the PP was preferably incorporated as a VP modifier. The following is a set of example test items used in a self-paced reading study.

- (11) a. The fireman / smashed down / the door / with the rusty lock / but smoke / overcame him.
- b. The fireman / smashed down / the door / with the heavy axe / but smoke / overcame him.
- c. The fireman / smashed down / a door / with a rusty lock / but smoke / overcame him.
- d. The fireman / smashed down / a door / with a heavy axe / but smoke / overcame him.

The reading time results showed that the PP region beginning in *with* was generally read faster when the DP containing *door* was indefinite as opposed to definite. The interaction of the definiteness of the DP and the preferred attachment of the PP was very nearly significant ($p = .056$). Spivey-Knowlton & Sedivy attributed this effect to a referential constraint that leads the parser to an analysis of the DP including *door* that allows for the simplest discourse model along the lines of PARSIMONY (Crain & Steedman 1985) and REFERENTIAL SUPPORT (Altmann & Steedman 1988).

The parser is facing this same situation in the cases we are concerned with involving BIAC attachment. A decision must be made between nominal-modification (the IRC reading) or verbal-modification (the RatC or PC readings). Thus, I will take the findings above to be an insight into the parser's strategy for incorporating a BIAC into the root clause. However, I will remain agnostic as to the exact force behind this behavior and simply take it at face value as an effect of the definiteness of the potential Head. With an indefinite DP, then the BIAC will be incorporated into the nominal domain as a modifier of the NP. If, on the other hand, the potential Head is definite, then the BIAC will be incorporated into the verbal domain.

The chosen domain of modification will indirectly determine whether the parser postulates an NOS in the BIAC. Given that the only available parse provided by the grammar in the case of nominal modification is an IRC, then an NOS must be posited early and constructed actively and predictively. Committing to a verbal modification analysis will make available multiple parses, some that contain an NOS (RatC or SPC) and some that do

not (OCP). The MCP predicts that the parser will pursue either the RatC or SPC parse¹³ and avoid the postulation of a chain.

The following experiment has been designed so as to be sensitive to the parser's preference for finding a gap or an overt argument in the embedded internal argument position and whether or not this preference significantly interacts with the definiteness of the DP preceding the BIAC. Therefore, we will be concerned with identifying the processing signatures of a predictive search when the assigned Head is indefinite and a lack of these signatures of a predictive search when the assigned Head is definite.

3.1 Methods

Participants

44 native American English speakers were recruited from the undergraduate population at the University of Massachusetts Amherst. All were compensated with course credit and were naive to the purpose of the experiment.

Materials

24 quadruples like that shown below in (12) were counterbalanced across 4 lists, which included 96 fillers. The conditions pivoted on two factors: the definiteness of the *potential Head*, which was the DP adjacent to the BIAC, and the presence (ARG condition) or absence (GAP condition) of an internal argument inside the BIAC.

- (12) a. INDEF(INITE)-GAP
 The landlord / hired *a lawyer* / to talk to / about the building code violations / discovered last month.
- b. INDEF(INITE)-ARG(UMENT)
 The landlord / hired *a lawyer* / to talk to / her tenants / about the

¹³ The choice between pursuing either the RatC or SPC reading comes down to the parser's preferred resolution of PRO in the BIAC. In Overfelt (in progress), I argue that the parser prefers to assign as the controller for PRO the structurally lowest available argument that could suitably act as the agent of an event. Thus, when the matrix subject and internal argument are both a human-animate, the internal argument is taken as the controller for PRO. This asymmetry is unexpected under a memory retrieval model of reactivation (cf. McElree 2000) for reasons concerning the information-structural prominence of the arguments. Thus, I suggest that a backwards search is being performed by the parsing mechanism. This search process I suggest can be conceived of in the same light as the backward search process proposed in Dillon et al. (in preparation) for the identification of an antecedent for long-distance reflexives in Mandarin.

building code violations / discovered last month.

c. DEF(INITE)-GAP

The landlord / hired *the lawyer* / to talk to / about the building code violations / discovered last month.

d. DEF(INITE)-ARG(UMENT)

The landlord / hired / *the lawyer* / to talk to / her tenants / about the building code violations / discovered last month.

The definiteness of the potential Head was encoded with the determiner *the* while indefiniteness was always indicated with the quantifier *a*. When the region pivoting on the gap/argument alternation contained an overt argument, a two word DP was always chosen that would be compatible with the scenario being described. A gap in this region was meant to be a plausible place to integrate the content of the potential Head. The embedded verb was chosen so as to be strongly preferred transitive or obligatorily transitive to ensure the presence of an effect indicating whether or not the parser was engaged in an active gap-filling procedure. Verbs were chosen based on verb bias counts from Frazier & Clifton Jr. (1989), Staub et al. (2006), and Gahl et al. (2004). Alternatively, a particle immediately followed the verb in the same presentation region. The regions consisted of between 1-5 words but averaged between 2-3. Each item contained two adjunct clauses to ensure that the critical region never appeared sentence final.

Procedure

The materials were presented region-by-region in a non-cumulative self-paced reading task using the Linger software¹⁴ on a MacBook Pro laptop with built-in keyboard. Each item initially appeared as a row of dashes, which were replaced left-to-right region-by-region at the push of the space bar. Every item was followed by either an acceptability rating question or a comprehension question to ensure that participants were attending to the materials. Before the experiment began the participants were exposed to several practice items to familiarize them with a self-paced reading task and introduce them to the types of questions they would be asked. The experiment took approximately 30 minutes.

¹⁴ (Written by Doug Rhode at MIT: <http://tedlab.mit.edu/dr/Linger/>)

3.2 Empirical predictions

If the parser engages in an active search for a gap position within the BIAC when the potential Head is indefinite, then we should see significantly longer reading times in the INDEF-ARG condition than in the DEF-ARG condition at the DP region. This will be indicative of a filled-gap effect, first identified in Stowe (1986). This effect is the increase in reading time that results when the parser experiences processing difficulty upon encountering a potential gap site in which to integrate a WH-element and this position is occupied by an overt argument. In the cases at hand, we will be seeing the effect of the parser initially positing an IRC parse in the INDEF condition and having to do a reanalysis when it reaches the argument, but an SPC or RatC parse being posited in the DEF condition and this analysis being confirmed at the embedded internal argument position.

Similarly, we might also expect longer reading times in the INDEF-ARG condition when compared to the INDEF-GAP condition. When the parser finds an overt internal argument in the BIAC, it will have disconfirming evidence for the IRC parse and be forced to reanalyze the structure. This late filled-gap effect should be visible as significantly longer reading times in the INDEF-ARG condition in the adjunct regions.

The current hypothesis also predicts a processing disruption leading to increased reading times in the DEF-GAP condition as compared to the INDEF-GAP and DEF-ARG conditions. These will be cases in which the parser was not actively constructing a dependency chain and so should be surprised to encounter an obligatorily transitive verb with a missing argument. This would suggest that the SPC or RatC parse is initially pursued in the DEF condition and then disconfirmed when the gap is encountered. I will refer to this effect as an *unpredicted-gap effect*. As noted, this effect should only surface with strongly preferred transitive verbs and obligatorily transitive verbs.

Finally, we should find that there is an interaction between the definiteness of the potential Head and the presence or absence of an argument in the embedded internal argument position. The presence of an interaction will be simulated by computing the F1 statistic through comparing the differences between the sum of the means of the cases that are predicted to involve reanalysis (DEF-GAP and INDEF-ARG) and those that are not predicted to involve reanalysis (DEF-ARG and INDEF-GAP).

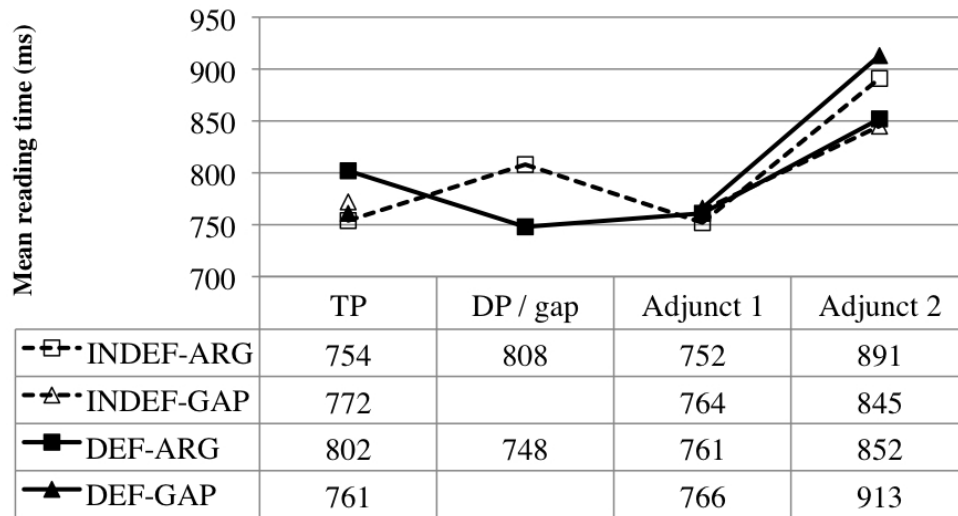


Figure 1: Mean reading times for Experiment 1

3.3 Results

The region-by-region mean reading times (ms) for each of the four conditions are presented in Figure 1 on the following page. The results of paired t-tests comparing the differences between the means of the conditions for which predictions were made are provided in Table 1 below. In the TP region (*to talk to*), there were no significant differences in the means or a difference in means suggesting an interaction of the conditions. Thus, there doesn't appear to be any sign of processing difficulty encountered before the BIAC is actually received.

In the DP region (*e₁ / her tenants*) there was a very near significant difference between the means in the INDEF-ARG condition and the DEF-ARG condition ($t_{(43)} = 1.87$, $p = 0.068$). The longer reading times in the INDEF-ARG condition suggests that the parser experienced processing difficulty when an overt argument occupied the embedded internal argument position and the potential Head was indefinite. The shorter reading times for the DEF-ARG condition suggest that the parser did not experience processing difficulty when it found an overt argument in the BIAC when the potential Head was definite.

There were no significant differences between means or a difference in the means suggesting an interaction in the first adjunct region (*about the*

Prediction	TP	DP	Adjunct 1	Adjunct 2
IDEF-ARG	$t_{(43)} = 1.09$	$t_{(43)} = 1.87$	$t_{(43)} = 0.15$	$t_{(43)} = 0.71$
> DEF-ARG	$p = 0.28$	$p = 0.068$	$p = 0.88$	$p = 0.48$
	95% CI =	95% CI =	95% CI =	95% CI =
	-40.0 to 135	-4.71 to 126	-116 to 135	-71 to 150
DEF-GAP	$t_{(43)} = 0.30$	—	$t_{(43)} = 0.04$	$t_{(43)} = 1.53$
> INDEF-GAP	$p = 0.76$	—	$p = 0.97$	$p = 0.13$
	95% CI =	—	95% CI =	95% CI =
	-64.1 to 86.9		-90.2 to 94	-21.4 to 158
INDEF-ARG	$t_{(43)} = 0.52$	—	$t_{(43)} = 0.28$	$t_{(43)} = 0.97$
> INDEF-GAP	$p = 0.76$	—	$p = 0.78$	$p = 0.34$
	95% CI =	—	95% CI =	95% CI =
	-51.6 to 87.6		-76.0 to 101	-1.76 to 125
DEF-GAP	$t_{(43)} = 1.18$	—	$t_{(43)} = 0.11$	$t_{(43)} = 1.96$
> DEF-ARG	$p = 0.25$	—	$p = 0.91$	$p = 0.056$
	95% CI =	—	95% CI =	95% CI =
	-29.3 to 111		-91.1 to 101	-1.76 to 125
INTERACTION	$t_{(43)} = 1.07$	—	$t_{(43)} = 0.10$	$t_{(43)} = 1.78$
(F1)	$p = 0.29$	—	$p = 0.92$	$p = 0.08$
	95% CI =	—	95% CI =	95% CI =
	-51.8 to 169		-142 to 156	-13.9 to 228

Table 1: Experiment 1 reading time results

building code violations). Instead, we saw in the second adjunct region (*discovered last month*) a very near significant difference between the DEF-GAP and DEF-ARG conditions ($t_{(43)} = 1.96$, $p = 0.056$). This suggests that the parser experienced a disruption in processing when a gap was found after the strongly/obligatorily transitive predicate in the BIAC when the potential Head was definite. Again, the shorter reading times in the DEF-ARG condition suggest that the parser did not experience a disruption when an overt argument was identified inside the BIAC.

We also saw a marginal effect in the simulation of an interaction between the definiteness of the Head and the presence of an argument ($t_{(43)} = 1.78$, $p = 0.08$) in the second adjunct region. I will take this to be indicative

of a tendency for the parser to only continue to pursue the IRC parse when the potential Head is indefinite.

3.4 Discussion of Experiment 1

The predictions were borne out in longer reading times at the embedded internal argument position when the preceding DP was indefinite as compared to when it was definite. Encountering the overt argument in the indefinite condition resulted in the filled-gap effect we were looking for. When the Head was definite, there was no disruption in processing upon encountering an overt argument and so no observable penalty in reading times. I interpret this to mean that the parser was actively constructing an NOS in the INDEF condition but not the DEF condition.

We also saw that encountering a gap in the BIAC would induce processing difficulty when the potential Head was definite. This effect showed up in the second adjunct region, which was read slower in the GAP condition. This can be interpreted as the unpredicted-gap effect expected to arise when the parser wasn't actively constructing a syntactic chain. When the required argument was present, as in the DEF-ARG condition however, then there was not a disruption in processing and therefore no slow down in the participants' rate of reading. This suggests that the parser did not pursue a parse that included an NOS given a definite potential Head.

Why it is that this effect was realized on the second adjunct as opposed to the first adjunct is not entirely clear. This is most plausibly related to the fact that an adjunct clause is not as definitively disambiguating as a DP. When the participants encountered a PP when they expected to find a DP in the DEF-GAP condition, it may have been the case that they suspected that the DP might simply be delayed. For instance, the parser may have been holding out for a possible Heavy-NP Shift structure. This is expected if participants gave up on finding the argument later in the input and reanalyzed the structure to one with an NOS only once the second adjunct (or the end of the sentence) was received.

The difference between the means of the INDEF-ARG and INDEF-GAP conditions was not significant contra the predictions made above. The most obvious and least informative answer might be that that reanalysis is simply easier in the indefinite cases as compared to the definite cases. This could be tied into the analysis one chooses to give to the means and cost of

updating the discourse analysis. It may simply be more difficult to update a discourse model that accommodates more unsupported presuppositions (cf. Crain & Steedman 1985; Altmann & Steedman 1988) as would be the case in the DEF conditions.

Overall the results of this experiment support the hypothesis that the postulation of a null-operator is avoided unless the grammar otherwise requires the parser to posit one early in the parse.¹⁵ We have also seen that indefinite DPs lead to a parse with an NOS more so than do definite DPs. From this, I suggest that an indefinite DP leads the parser to IRC parse when all else is equal. A definite DP on the other hand leads the parser to either a RatC or SPC parse when all else is equal. The definiteness of the DP leads the parser to associate the BIAC into the verbal domain. The MCP essentially “ranks” the three available parses at this point, dispreferring the analysis that contains an NOS.

4. Experiment 2

The results of Experiment 1 have lead us to expect that either a RatC or an SPC parse will be entertained before an OPC interpretation of the BIAC. Essentially, once the IRC parse has been taken out of the equation, sufficient bottom-up grammatical evidence will be necessary to establish an NOS. Therefore, this experiment too has been designed to be sensitive to the parser’s preferences for finding a gap in the embedded internal argument position, but we limit the potential Head to definite DPs.

As we’ll see the experimental factors pivot on the presence of an overt argument in the BIAC and the transitivity preference of the embedded predicate. Alternating the transitivity preference of the embedded predicate will provide a way of testing for the unpredicted-gap effect in Experiment 1. If the parser is able to accommodate a preferred transitive verb in its intransitive frame, then we won’t see a disruption upon finding a gap. This will be because the parser has overlooked this analysis. This in turn will sug-

¹⁵ One draw back of this experiment is that we can’t be sure whether the parser, when reanalysis is necessary, has gone back to the IRC parse, or if it has moved on to an OPC parse. This is not an issue that is dealt with explicitly in this paper, however it is more of an issue that relates to the retrospective operation that constructs NOSs non-predictively instead of an issue directly related to the first-pass parsing preferences of null-operator structures. Thus, any further discussion of this issue will have to remain for future research.

gest that the parser is avoiding the postulation of a syntactic chain given the availability of an alternative parse, namely a RatC or SPC parse.

In addition, Pickering & Traxler (2003) found there that in the course of parsing a relative clause the parser will attempt to resolve a dependency chain by assigning the filler to a verb for which it is an implausible argument. Consider the pair of sentences below.

- (13) *Plausible*
That's the plane that the pilot / landed carefully / behind in the fog / at the airport.
- (14) *Implausible*
That's the truck that the pilot / landed carefully / behind in the fog / at the airport.

Pickering & Traxler found that it took longer for participants to read the region containing the embedded verb *landed* in the implausible condition than in the plausible condition. This indicates that the participants were integrating the RC Head as the filler of a gap following *landed* regardless of the fact that *landed* preferably selects PPs as opposed to NPs. Only in this way could the chain be evaluated and found to be semantically odd. The conclusion is that the subcategorization preference of a verb is not taken into account before the parser attempts to assign a filler to an A-position.

The reasoning here is that a filled-gap effect, if present, will be equally robust in both the preferred transitive and preferred intransitive conditions. This is exactly because the parser will be ignoring the subcategorization preference of the verb. Returning to the experiment at hand, any effect related to the valency preference of the verb will suggest that the parser is not engaged in an active chain resolution strategy.

4.1 Methods

Participants

26 native American English speaking participants were recruited from the undergraduate population at the University of Massachusetts Amherst. Participants were compensated with course credit and were naive to the purpose of the study.

Materials

24 pairs (12 with preferred transitive verbs and 12 with preferred intransitive verbs) were counterbalanced across 2 lists including 48 fillers. The two conditions pivoted on the presence or absence of an embedded argument and the transitivity bias of the *embedded verb*.

- (15) a. TR(ANSITIVE)-GAP
Our classmates / borrowed my notes / to *study* / for the exam / next week.
- b. TR(ANSITIVE)-ARG(UMENT)
Our classmates / borrowed my notes / to *study* them / for the exam / next week.
- c. INTR(ANSITIVE)-GAP
The businessman /packed up his things / to *move* / to a new office / down the hall.
- d. INTR(ANSITIVE)-ARG(UMENT)
The businessman /packed up his things / to *move* them / to a new office / down the hall.

The transitivity of the embedded verb was determined based on verb bias counts in Staub et al. (2006) and Gahl et al. (2004). Both the preferred transitive and preferred intransitive verbs were chosen based on the plausibility of use in their alternative valency frame. When an internal argument was present, it appeared as a pronoun intended to be coreferent with the DP adjacent to the BIAC. Each experimental item contained 5 regions that consisted of between 2-4 words. Two adjuncts were included so as to provide at least one spillover region for each of the ARG and GAP conditions. The DP to which the BIAC was adjacent was always definite in order to discourage the initial IRC parse.¹⁶

Procedure

The non-cumulative phrase-by-phrase SPR task used in Experiment 1 was repeated here with the same equipment. However, in this experiment, each item was followed by a comprehension question that asked the participants to choose between two paraphrases for the sentence they had just read. This

¹⁶ It was decided that full disambiguation by means of having the BIAC non-adjacent to the potential binder for the gap could unfairly bias against the construction of the NOS.

was done in an attempt to determine the interpretation the parser assigned to the sentence (RatC or OPC). Given the subtlety between the differences in the meanings of some of these structures and various experimental errors, the results were inconclusive. Therefore, this part of the experiment is not presented here. The experiment took approximately 30 minutes.

4.2 Empirical predictions

At the region including the embedded internal argument position (*to move/study them*) we expect that the ARG conditions are both read slower than the GAP conditions simply due to the fact that the former contain a pronoun. However, if a filled-gap effect is also contributing to the reading time difference between these two conditions, then we wouldn't predict a significant difference between the means of the TR-GAP and TR-ARG conditions or between the INTR-GAP and INTR-ARG conditions. We expect this given the findings mentioned previously in Pickering & Traxler (2003).

On the other hand, if we were to see an effect of the subcategorization preference of the verb, we might expect to find this manifested as a significant difference between the TR-ARG and INTR-ARG conditions at the TP region. This would suggest first that the parser has chosen to first pursue an analysis that does not require the resolution of a long-distance dependency and would be attending to the subcategorization preferences of the predicates. However, we wouldn't necessarily expect see a difference between the TR-GAP and INTR-GAP conditions. The lack of a significant difference between the GAP conditions in the TP region or the first spillover region would also suggest that the parser has entirely overlooked the OPC parse on the first-pass through the string. Without an obligatorily transitive verb, the parser will easily be able to accommodate the transitive verb in its intransitive frame.

Of course, it is possible that the parser will be able to identify the availability of the OPC interpretation. If this is the case, then we should see a reading time slowdown for the TR-GAP condition but not the TR-ARG condition in the second Adjunct region. This will indicate a disruption at the potential gap position and can be taken to signal the need for the parser to perform a reanalysis to insert the NOS.

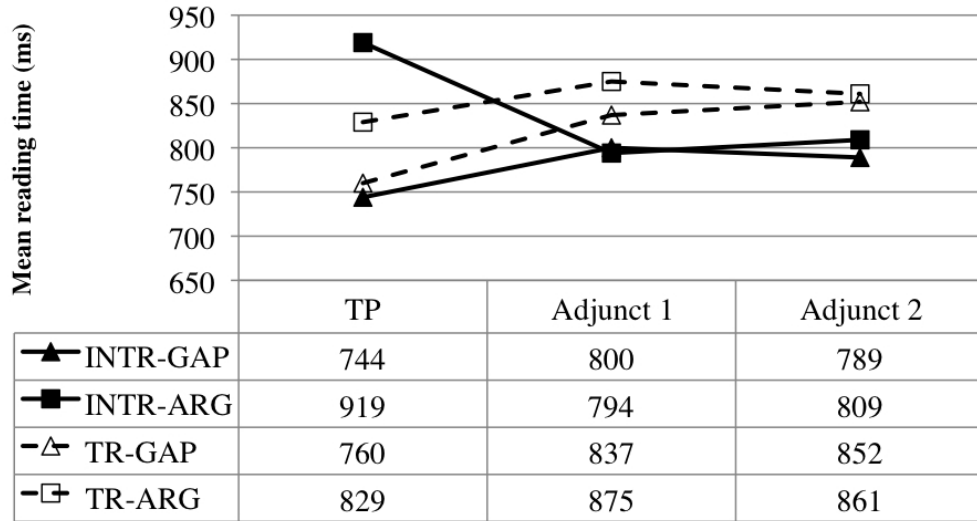


Figure 2: Mean reading times for Experiment 2

4.3 Results for Experiment 2

The region-by-region mean reading times (ms) for each of the four conditions are presented in Figure 2. In Table 2 on the following page are the results of paired t-tests comparing the differences between the means of the conditions for which predictions were made. In the TP region (*to move/study e₁/them*) there was a near significant difference between the means in the INTR-ARG and TR-ARG conditions ($t_{(25)} = 2.00$, $p = 0.057$). The longer reading times in the INTR-ARG condition suggest that the parser experienced processing difficulty when the overt argument followed an intransitive verb but not a transitive verb. We also see that there is no significant difference between the means of the TR-GAP and INTR-GAP conditions ($t_{(25)} = 0.41$, $p = 0.65$) suggesting that the parser did not experience any processing difficulty when it encountered a transitive verb that did not have an overt argument.

There were no significant differences between the means of any conditions in either of the adjunct regions. This suggests that the parser did not experience any processing difficulty in either of the GAP conditions after finding that the embedded verb did not have an overt argument. Finally, the simulated F1 statistic does not show a significant interaction between the transitivity bias of the embedded predicate and the presence or absence of

Prediction	TP	Adjunct 1	Adjunct 2
INTR-ARG = INTR-GAP	— — —	$t_{(23)} = 0.12$ $p = 0.88$ 95% CI = 97.9 to 110	$t_{(43)} = 0.46$ $p = 0.48$ 95% CI = -68.6 to 109
INTR-ARG = TR-ARG	$t_{(23)} = 2.00$ $p = 0.057$ 95% CI = -2.78 to 182	$t_{(23)} = 1.32$ $p = 0.20$ 95% CI = -45.4 to 208	$t_{(23)} = 0.98$ $p = 0.34$ 95% CI = -57.4 to 162
TR-GAP = INTR-GAP	$t_{(23)} = 0.41$ $p = 0.65$ 95% CI = -67.0 to 101	$t_{(23)} = 0.58$ $p = 0.57$ 95% CI = -95.5 to 170	$t_{(23)} = 0.89$ $p = 0.38$ 95% CI = -82.5 to 209
TR-ARG = TR-GAP	— — —	$t_{(23)} = 0.84$ $p = 0.41$ 95% CI = -55.3 to 131	$t_{(43)} = 0.15$ $p = 0.88$ 95% CI = -113 to 131
INTERACTION (F1)	$t_{(23)} = 1.68$ $p = 0.11$ 95% CI = -24.4 to 238	$t_{(23)} = 0.57$ $p = 0.57$ 95% CI = -113 to 201	$t_{(23)} = 0.15$ $p = 0.88$ 95% CI = -137 to 159

Table 2: Experiment 2 reading time results

an overt argument. This suggests that the parser did not categorically experience disruptions in processing due the valency of the embedded verb not being met.

4.4 Discussion of Experiment 2

The hypothesis that participants were expected to overlook the OPC entirely on the first-pass parse is supported by the data above. The presence of a very near significant difference between the reading times of the TR-ARG and INTR-ARG conditions suggest that the parser did not experience a filled-gap effect upon encountering the embedded argument. Recall that according to the work done by Pickering & Traxler (2003) such an effect would have been independent of the transitivity bias of the predicate. The

absence of a filled-gap effect and, more importantly, the presence of an effect of transitivity bias can be taken as evidence that the parser was not engaged in a predictive parsing strategy to resolve a long-distance dependency chain.

We also saw no significant difference between the mean reading times in the TR-GAP and INTR-GAP conditions in any region. This null-effect is in line with the prediction that the parser would be able to accommodate a preferred transitive verb in its intransitive frame. Without an obligatorily transitive verb, the grammar provides the parser with an escape hatch through which it may avoid postulating an NOS.

This does, however, raise the question of why there should be an asymmetry in the effects of the subcategorization preference of the verb. Obviously we have to say that it is somehow easier to interpret a preferred transitive verb in its intransitive frame than it is to interpret a preferred intransitive verb in its transitive frame (i.e., TR-GAP is an easier accommodation than INTR-ARG). This might simply be tied to the syntactic parsing principles (Frazier 1978), which prefer building as little structure as possible. This could in turn make the absence of an argument easily accommodated as phrase structure is already being minimized. When extra structure is necessary however, say to accommodate an unpredicted argument, the processing procedure is disrupted while the extra argument structure is incorporated.

Together these results suggest that the parser chose to initially pursue an analysis that did not require the postulation of a potentially unnecessary syntactic chain. Furthermore, we saw that the parser was able to interpret an embedded preferred transitive verb in its intransitive frame. This suggests that the parser regularly overlooked the availability of the OPC parse and thus the possibility for inserting an NOS. This can be taken as support for the hypothesis that syntactic chains are avoided by the parser and only postulated (either early or late in the parsing procedure) when they are required by the grammar.

5. Conclusion

The results of the experiments above, while interesting, are not terribly surprising if it's true that the parser typically does not like to postulate empty categories for which it has no evidence (De Vincenzi 1991). What is sur-

prising is the observation that the human sentence-processor has at its employment multiple strategies by which to resolve the same long-distance dependency. As we saw, the active or bottom-up resolution of an NOS is dependent on when in the parsing procedure the chain is identified.

The necessary conclusion to be reached from these findings is that the resolution strategy employed in interpreting an NOS cannot be dependent on any language particular properties concerning how the chain is represented in the string presented to the parser. Thus, the first-resort and last-resort nature of dependencies in general is an illusory effect of the options made available to the parser by the grammar at the points in the parsing procedure that the chain is realized. This means for English WH-elements that they do not themselves trigger a predictive parsing strategy. Instead, given the grammar, when the parser encounters a moved WH-element it will recognize the presence of a syntactic chain and will resolve it actively and predictively.

A number of other studies have been conducted which, when assembled, provide further evidence that the processing strategy employed in the resolution of a long-distance dependency is not predictable based on the representation of the chain. In recent work on Heavy-NP Shift, Staub et al. (2006) found that the parser interpreted a preferred transitive verb missing an argument in its intransitive frame before engaging in a search for a rightward displaced DP. However, when the gap followed an obligatorily transitive predicate, the data suggest that the parser in fact did engage in an active search for a DP to establish as the filler of a movement chain. For our purposes, this finding suggests that the nature of the first link of the movement chain that the parser encounters does not condition the type of resolution strategy employed.

Sprouse et al. (2011) and Xiang et al. (2010) demonstrated for English and Mandarin respectively that the parser will employ a retrospective resolution strategy for a WH-dependency when the chain is necessarily contained in the previously parsed material. Alternatively, Miyamoto & Takahashi (2000), Ueno & Kluender (2003), and Aoshima et al. (2004) have demonstrated for Japanese that the parser engages in a predictive parsing strategy after finding an in-situ WH-element in order to locate the scope-marking particle associated with it. Miyamoto & Takahashi found that the parser experienced a disruption in processing when the first verb encoun-

tered after the in-situ WH-element carried a declarative particle as opposed to a question particle. Assuming that the particle is base-generated, then these facts suggest that an active search strategy is not necessarily initiated only by ex-situ constituents but by in-situ constituents as well.

If this claim turns out to capture the actual state of affairs concerning long-distance dependency resolution, then we have further evidence for suggesting that language processing is influenced directly by the grammar of a given language. And if the grammar is able to directly influence parsing as I have suggested here, then we blur the line slightly more between the competence and performance components of the language faculty. We might begin to think of these components not as separate cognitive systems of the language faculty, but simply as two observable manifestations of one and the same system.

References

- Alphonse, Carl & Henry Davis (1997). Motivating non-directional movement. Beermann, Dorothee, David LeBlanc & Henk van Riemsdijk (eds.), *Rightward Movement*, John Benjamins, Amsterdam/Philadelphia, pp. 7–35.
- Altmann, Gerry & Mark Steedman (1988). Interaction with context during human sentence processing. *Cognition* 30, pp. 191–238.
- Aoshima, Sachiko, Colin Phillips & Amy Weinberg (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language* 51, pp. 23–54.
- Bader, Markus & Michael Meng (1999). Subject-object ambiguities in german embedded clauses: An across the board comparison. *Journal of Psycholinguistic Research* 28, pp. 121–141.
- Bhatt, Rajesh (2006). *Covert modality in non-finite contexts*. Mouton de Gruyter, Philadelphia, PA.
- Browning, Marguerite (1987). *Null Operator Constructions*. Ph.D. thesis, MIT, Cambridge, MA.
- Chomsky, Noam (1980). On binding. *Linguistic Inquiry* 11, pp. 1–46.
- Chomsky, Noam (1986). *Barriers*. The MIT Press, Cambridge, MA.
- Crain, Stephen & Janet Dean Fodor (1985). How can grammars help

- parsers? Dowty, David R., Lauri Karttunen & Arnold M. Zwicky (eds.), *Natural Language Parsing: Psycholinguistic, Computational, and Theoretical Perspectives*, Cambridge University Press, Cambridge, UK, pp. 94–128.
- Crain, Stephen & Mark Steedman (1985). On not being led up the garden path: The use of context by the psychological parser. Dowty, David R., Lauri Karttunen & Arnold M. Zwicky (eds.), *Natural Language Parsing: Psycholinguistic, Computational, and Theoretical Perspectives*, Cambridge University Press, Cambridge, UK.
- De Vincenzi, Marica (1991). *Syntactic parsing strategies in Italian: The Minimal Chain Principle*. Kluwer.
- Dillon, Brian, Wing Yee Chow, Matthew Wagers, Taomei Guo, Fengqin Liu & Colin Phillips (in preparation). *The structure-sensitivity of memory access: Evidence from Mandarin Chinese*. Ms.
- Engdahl, Elisabet (1983). Parasitic gaps. *Linguistics and Philosophy* 6, pp. 5–34.
- Faraci, Robert A. (1974). *Aspects of the grammar of infinitives and for-phrases*. Ph.D. thesis, MIT, Cambridge, MA.
- von Fintel, Kai & Sabine Iatridou (2005). *What to do if you want to go to Harlem: Anankastic conditionals and related matters*. Ms., Cambridge, MA.
- Fodor, Janet Dean (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry* 9, pp. 427–473.
- Frazier, Lyn (1978). *On comprehending sentences: Syntactic parsing strategies*. Ph.D. thesis, University of Connecticut, Storrs, CT.
- Frazier, Lyn (1987). Syntactic processing: Evidence from dutch. *Natural Language & Linguistic Theory* 5, pp. 519–559.
- Frazier, Lyn & Charles Clifton Jr. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4, pp. 93–126.
- Frazier, Lyn & Charles Clifton Jr. (1996). *Construal*. The MIT Press.
- Frazier, Lyn & G.B. Flores D’Arcais (1989). Filler-driven parsing: A study of gap filling in dutch. *Journal of Memory and Language* 28, pp. 331–344.
- Gahl, Susanne, Dan Jurafsky & Douglas Roland (2004). Verb subcategorization frequencies: American english corpus data, methodological

- studies, and cross-corpus comparisons. *Behavior Research Methods, Instruments, & Computers* 36, pp. 432–443.
- Gibson, Edward (1998). Syntactic complexity: Locality of syntactic dependencies. *Cognition* 68, pp. 1–76.
- Hackl, Martin & Jon Nissenbaum (accepted). A modal ambiguity in *for*-infinitival relative clauses. *Natural Language Semantics* .
- Huettner, Alison (1989). *Adjunct infinitives in English*. Ph.D. thesis, UMass, Amherst, MA.
- Johnson, Kyle (1991). Object positions. *Natural Language & Linguistic Theory* 9, pp. 577–636.
- Jones, Charles (1991). *Purpose clauses: Syntax, thematics, and semantics of English purpose constructions*. Kluwer, Dordrecht.
- Kayne, Richard (1983). Connectedness. *Linguistic Inquiry* 14:2, pp. 223–249.
- Kayne, Richard (1994). *Antisymmetry*. The MIT Press, Cambridge, MA.
- Kurtzman, Howard & Loren Crawford (1991). Processing parasitic gaps. Sherer, Tim (ed.), *Proceedings of NELS 21*, UMass GLSA, Amherst, MA, pp. 217–231.
- Maia, Marcus (1997). The processing of object anaphora in brazilian portuguese. *Researches Linguistiques de Vincennes* 26, pp. 151–172.
- McElree, Brian (2000). Sentence comprehension is mediated by content-adressable memory structures. *Journal of Psycholinguistic Research* 29, pp. 111–123.
- Miyamoto, Edson & Shoichi Takahashi (2000). The processing of wh-phrases and ininterrogative complementizers in japanese. Akatsuka, Noriko, Susan Strauss & Bernard Comrie (eds.), *Japanese/Korean Linguistics*, vol. 10, pp. 62–75.
- Nissenbaum, Jon (2000). *Investigations of covert phrase movement*. Ph.D. thesis, MIT, Cambridge, MA.
- Nissenbaum, Jon (2005). Sates, events and vp structure: Evidence from purposive adjuncts. NELS 36, UMass, Amherst, MA.
- Overfelt, Jason (2010). *The syntax and semantics of infinitival adjunct clauses*. Ms., University of Massachusetts, Amherst, MA.
- Overfelt, Jason (2011). *Parsing preposed DPs and the role of information structure*. Ms., University of Massachusetts, Amherst, MA.

- Overfelt, Jason (in progress). *Resolving PRO in bare infinitival adjunct clauses*. Ms., University of Massachusetts, Amherst, MA.
- Pablos, Leticia (2006). *Pre-verbal structure building in Romance languages and Basque*. Ph.D. thesis, University of Maryland, College Park, MD.
- Pickering, Martin (1994). Processing local and unbounded dependencies: A unified account. *Journal of Psycholinguistic Research* 23, pp. 323–352.
- Pickering, Martin & Guy Barry (1991). Sentence processing without empty categories. *Language and Cognitive Processes* 6, pp. 229–259.
- Pickering, Martin & Matthew J. Traxler (2003). Evidence against the use of subcategorisation frequency in the processing of unbounded dependencies. *Language and Cognitive Processes* 18, pp. 469–503.
- Ross, John Robert (1967). *Constraints on variables in syntax*. Ph.D. thesis, MIT, Cambridge, MA.
- Seely, Daniel (1987). The dependency hypothesis: Toward a theory of the processing of parasitic gaps. *Proceedings of the 6th WCCFL*, University of Arizona.
- Spivey-Knowlton, Michael & Julie Sedivy (1995). Resolving attachment ambiguities with multiple constraints. *Cognition* 55, pp. 227–267.
- Sprouse, Jon, Shin Fukuda, Hajime Ono & Robert Kluender (2011). Reverse island effects and the backward search for a licensor in multiple *wh*-questions. *Syntax* 14, pp. 179–203.
- Staub, Adrian, Charles Clifton Jr. & Lyn Frazier (2006). Heavy-np shift is the parser's last resort: Evidence from eye movements. *Journal of Memory and Language* 54, pp. 389–406.
- Stowe, Laurie (1986). Parsing *wh*-constructions: Evidence for on-line gap locations. *Language and Cognitive Processes* 1, pp. 227–245.
- Ueno, Mieko & Robert Kluender (2003). On the processing of Japanese *Wh*-questions: Relating grammar and brain. Garding, G. & M. Tsujimura (eds.), *Proceedings of the 22nd WCCFL*, Cascadilla, Somerville, MA, pp. 491–504.
- Vergnaud, Jean-Roger (1974). *French relative clauses*. Ph.D. thesis, MIT, Cambridge, MA.
- Wagers, Matthew & Colin Phillips (2009). Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguis-*

tics 45, pp. 395–433.

Whelpton, Matthew (2002). Locality and control with infinitives of result.
Natural Language Semantics 10, pp. 167–210.

Xiang, Ming, Brian Dillon, Matthew Wagers, Fengqin Liu & Taomei Guo
(2010). Processing wh-movement dependencies in a language without wh-movement. Poster presented at CUNY Conference on Sentence Processing.

Processing effects of an indeterminate future: Evidence from self-paced reading

Adrian Staub and Charles Clifton, Jr.

University of Massachusetts Amherst

Consider statements such as (1a-b):

- (1) a. Sally ate an apple before lunch.
- b. Sally will eat an apple before lunch.

The starting point of the present paper is the observation that (1a) and (1b) are likely to affect the beliefs of a listener or reader in qualitatively different ways. In the absence of conflicting evidence, (1a) is likely to generate the belief that Sally did, in fact, eat an apple before lunch. On the other hand, (1b) is likely to generate the belief that Sally will probably eat an apple before lunch, or that for all the speaker knows, Sally will eat an apple before lunch. In other words, (1a) may be regarded as true, while (1b) may be regarded as merely likely.

The different effects of (1a) and (1b) on a comprehender's set of beliefs may be understood in at least two ways. It may be understood in terms of naïve metaphysics: according to what is (arguably) a standard set of metaphysical assumptions, sentence (1a) has a determinate truth value at its time of utterance, while sentence (1b) does not. At the time (1b) is uttered, it is neither true nor false that Sally will eat an apple before lunch, as we generally assume some form of indeterminism about the future, which dictates that it cannot be completely settled, prior to Sally's eating the apple, whether she will do so or not. Perhaps (1b) acquires a truth-value later on, i.e., after lunch has begun (see, e.g., Belnap, Perloff, and Xu 2001, McFarlane 2003, for extended philosophical discussion). Alternately, it may be argued that both sentences do have a determinate truth-value at the time of utterance, but that as listeners, we recognize that statements about the past tend to have a stronger evidentiary base than statements about the future. An

assertion about the past may be based on direct observation, testimony from someone who has made a direct observation, etc.; but an assertion about the future is, at best, based on inductive inference from secure premises (e.g., Sally said that she will eat an apple before lunch; Sally has eaten an apple before lunch every day this month; there are in fact apples in the house).

In this paper we present an experiment designed to investigate on-line processing effects of the purported difference between the comprehender's attitude toward propositions expressed in the past tense and her attitude toward propositions expressed in the future tense. The experiment investigates whether a reader or listener becomes rapidly committed to the truth of a past-tense assertion like (1a), but not to the truth of a future-tense assertion like (1b). By rapidly, we mean within the time frame of processing the sentence itself.

The experiment is based on the following line of reasoning. If a comprehender takes a sentence such as (1a) to be true, but it then turns out that (1a) is merely the first clause of a disjunction, as in (2a) below, the comprehender will have to revise her beliefs to reflect the fact that (1a) may not be true. This belief revision may have a processing cost. But if the comprehender takes (1b) to be a prediction or statement of belief (i.e., not certainly true), then when it turns out that (1b) is merely the first clause of a disjunction, as in (2b) below, the processing cost associated with belief revision may be mitigated or absent altogether.

- (2) a. Sally ate an apple before lunch, or her mother gave her some carrots.
- b. Sally will eat an apple before lunch, or her mother will give her some carrots.

In order to assess the processing cost associated with belief revision in the two sentences, we compared reading times on sentences like (2a-b) with reading times on control sentences in which the word *either* was included at the beginning of the sentence, as in (3a-b), making clear to the reader early on that the initial clause is in fact a disjunct.

- (3) a. Either Sally ate an apple before lunch, or her mother gave her some carrots.
- b. Either Sally will eat an apple before lunch, or her mother will give her some carrots.

Thus, we are able to compare reading time on the critical material in (2a-b) with reading time on identical material in a context in which belief revision is unlikely. Any belief revision cost should take the form of inflated reading time on the second clause of the sentences in (2), compared to the corresponding sentences in (3).

1. Method

1.1 Participants

Forty-eight members of the University of Massachusetts community participated in the experiment, either for course credit or \$7. All were native speakers of English, and all were naive to the purpose of the experiment.

1.2 Materials

Twenty-four stimulus sets like that shown in (4) were constructed. Many of the items were based on the items used in Staub & Clifton (2006). This experiment demonstrated a general facilitation from the word *either* on reading of a second disjunct, whether the sentence involved clausal disjunction or noun phrase disjunction; in the Discussion section below we comment on the relationship between the present study and this earlier study. Unlike in the Staub & Clifton (2006) study, a comma was always included between clauses in order to prevent an initial noun phrase disjunction analysis. This incorrect analysis would also have been highly implausible, as the subject of the second clause was always an implausible object for the verb of the first clause.

- (4)
- | | |
|--|---------------------|
| a. Either John will borrow a rake, or his wife will buy one. | (either, future) |
| b. Either John borrowed a rake, or his wife bought one. | (either, past) |
| c. John will borrow a rake, or his wife will buy one. | (no either, future) |
| d. John borrowed a rake, or his wife bought one. | (no either, past) |

The 24 experimental sentences were separated into four lists, with one version of each stimulus set appearing on each list, and each list containing six sentences in each of the four experimental conditions. Each participant read the sentences on one of the four lists. The experimental sentences were intermixed with 114 unrelated filler sentences and presented in an individually randomized order to each participant, after two practice sentences. The full set of materials is presented in the Appendix.

1.3 Procedure

Participants were tested individually using a non-cumulative phrase-by-phrase self-paced reading paradigm (Just et al. 1982). Stimuli were presented and responses were collected on an IBM compatible computer running the E-Prime software package (Schneider et al. 2002). Each sentence was divided into four presentation regions, as shown below:

- (5) Either John / will borrow a rake, / or his wife / will buy one.

At the beginning of each trial, all characters were represented as dashes on the screen, with the first two regions on one line and the final two regions on the subsequent line. When the participant pressed the space bar, the dashes corresponding to the first region were replaced by characters. When the participant pressed the space bar again, these characters reverted to dashes, and the dashes in the second region were replaced by characters. This process continued until the participant finished reading the sentence and pressed the space bar for a final time, at which point the sentence was replaced by a two-choice comprehension question. The participant responded to the comprehension question by pressing a key. The duration of the experiment was approximately thirty minutes.

1.4 Results

The comprehension questions associated with the sentences in this experiment were answered correctly on 91% of trials, with no participant scoring below 75% correct.

The participant mean reading times for each region in each condition are displayed in Table 1. On region 1, reading times were longer when *either* was present, as would be expected based on the length difference ($F_1(1,47) = 202.05, p < .001$; $F_2(1,23) = 138.12, p < .001$). The effect of tense was not significant on region 1 ($F_1(1,47) = 2.59, p = .11$; $F_2(1,23) = 2.23, p = .15$), and the interaction effect was also not significant ($F_s < 1.5$). On region 2, reading times were longer for the future tense conditions than for the past tense conditions, which is again attributable to length differences ($F_1(1,47) = 20.93, p < .001$; $F_2(1,23) = 21.01, p < .001$). The effect of *either* was not significant ($F_1(1,47) = 1.76, p = .19$; $F_2(1,23) = 3.16, p = .09$), and the interaction effect was also not significant ($F_s < 1$). On region 3, reading times were 98 ms shorter when *either* was present than

when *either* was absent, a difference that was highly significant ($F_1(1, 47) = 36.26$, $p < .001$; $F_2(1, 23) = 30.94$, $p < .001$). This main effect of *either* was modulated by a significant interaction with tense: for the future tense conditions, the presence of *either* reduced reading time by 68 ms, while for the past tense conditions, the presence of *either* reduced reading time by 129 ms ($F_1(1, 47) = 7.48$, $p < .01$; $F_2(1, 23) = 6.08$, $p < .05$). The main effect of tense was not significant on region 3. On the final region, there was once again an effect of tense, with longer reading times in the future tense conditions due to length differences ($F_1(1, 47) = 11.82$, $p < .01$; $F_2(1, 23) = 13.07$, $p < .01$). No other effects approached significance on the final region (all F s < 1).

Condition	Region 1 (<i>Either</i>) John	Region 2 (<i>will borrow / borrowed</i>) a rake,	Region 3 <i>or his wife</i>	Region 4 (<i>will buy / bought</i>) one.
either, future	771	1015	691	951
either, past	810	911	662	865
no either, future	571	980	759	960
no either, past	583	894	791	896

Table 1: Participant mean reading times on each region, in milliseconds.

1.5 Discussion

We will confine our discussion to the results from region 3, as all other results are easily interpretable in terms of simple length differences. On region 3 reading times were significantly shorter when *either* was present, but this difference was significantly greater for the past tense sentences than for the future tense sentences.

The main effect of *either* obtained in the present experiment is a clear replication of one of the principal findings obtained by Staub & Clifton (2006), who used eyetracking to investigate the question of whether reading may be facilitated when a syntactic structure can be predicted in advance. In their experiment, participants read sentences containing clausal disjunction (which provided the basis for some of the materials in the present experiment), as well as sentences involving noun phrase disjunction (e.g., *The team took the subway or the train to get to the game.*) In addition, the word *either* could be present or absent before the disjunction. There were two

critical findings. First, reading time on the region consisting of the word or and the subsequent noun phrase was significantly reduced by the presence of *either*, as measured by both first pass time (the sum of the reader's eye fixation durations beginning with the first fixation on this region and continuing until the reader leaves the region for the first time, to the left or the right) and go-past time (the sum of the reader's eye fixation durations beginning with the first fixation on the region and continuing until the reader leaves the region to the right, i.e., including re-reading time). These effects were similar in size for the clausal disjunction and noun phrase disjunction sentences. Second, despite the fact that the clausal disjunction sentences were designed to render the incorrect noun phrase disjunction analysis highly implausible, there was some evidence of garden-pathing in these sentences when *either* was absent, in the form of an increase in regressive eye movements from the final region of the sentence. (Unlike in the present study, there was no comma at the clause boundary.) But when *either* was present, this garden path effect was eliminated entirely. On the basis of these results, Staub and Clifton argued that the word *either* enabled readers to predict the arrival of a disjunction, and that this prediction facilitated processing of the second disjunct when it ultimately arrived. Furthermore, the presence of *either* before a clausal disjunction eliminated any tendency to mistakenly analyze the structure as noun phrase disjunction (see also Staub, 2007, for additional evidence bearing on this question).

The significant interaction effect on region 3 suggests that when *either* was absent, readers did in fact become more strongly committed to the truth of a past-tense assertion in the first clause than to the truth of a future-tense assertion. The cost of belief revision is reflected in the fact that it took 129 ms longer for participants to read region 3 in the *no either, past* condition than in the *either, past* condition; we assume that in the latter condition, no belief revision was necessary. The corresponding difference was only 68 ms for the future tense conditions. This may be interpreted as suggesting that there is some belief-revision cost for the future tense, as well. We think it is more likely, however, that the 68 ms benefit when *either* is present is a syntactic predictability benefit, consistent with the conclusions of Staub & Clifton (2006).

2. Conclusion

The experiment presented here suggests that whether readers become committed to the truth of an assertion, while they are still reading the sentence, depends in part on tense. When the truth of a past tense assertion was called into question by the arrival of a disjoined second clause, there was significantly greater processing cost on this clause than when a future tense assertion was similarly called into question. We interpret this pattern as reflecting the fact that future tense assertions are likely to be regarded as only probably true; little if any belief revision is necessary if it becomes explicit, by means of disjunction, that a future tense assertion may or may not be true. As we mentioned in the introduction, the comprehender's tendency to withhold belief in the proposition expressed by a future tense sentence may be understood either in metaphysical terms (i.e., statements made at time t about events taking place at time $t + 1$ are neither true nor false at time t), or in epistemic terms (i.e., past tense assertions are generally based on stronger evidence than future tense assertions, and comprehenders recognize this).¹

On a more general note, this experiment is, to our knowledge, the first clear suggestion that belief revision has on-line processing costs that are distinct from other sources of processing difficulty such as syntactic revision. We think this may be a fertile area for future sentence and discourse processing research.

¹ Florian Schwarz has provided the following interesting example in support of the second account. Suppose an archeologist discovers an ancient skull with clear signs of damage. She reports, "He was killed in a fight or he fell from a cliff." Given the context, a listener is likely to recognize the somewhat shaky evidentiary base for the assertion in the first clause, despite the use of the past tense. This predicts that the disjunction should cause little or no processing difficulty associated with belief revision. Indeed, our own intuitions are that no processing difficulty is present in this case.

References

- Just, Marcel A., Patricia A. Carpenter & Jacqueline D. Wooley (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 111, pp. 228–238.
- Schneider, Walter, Amy Eschman & Anthony Zuccolotto (2002). *E-Prime User's Guide*. Psychology Software Tools, Inc., Pittsburgh, PA.
- Staub, Adrian (2007). The return of the repressed: Abandoned parses facilitate syntactic reanalysis. *Journal of Memory and Language* 57, pp. 299–323.
- Staub, Adrian & Charles Clifton, Jr. (2006). Syntactic prediction in language comprehension: Evidence from *Either...or*. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32, pp. 425–436.

Appendix

Experimental Materials. Future tense versions are shown below; past tense versions were constructed as in (4) above.

1. (Either) John will borrow a rake, or his wife will buy one.
2. (Either) Ms. Haywood will plan a picnic, or the kids will go swimming.
3. (Either) the Congress will pass the bill, or the President will issue an order.
4. (Either) Linda will buy the red car, or her husband will lease the green one.
5. (Either) Dr. Wendell will perform the surgery, or the nurse will make a splint.
6. (Either) Liza will sing her famous song, or the host will dance with a bear.
7. (Either) Rudolph will pull the sleigh, or the leader will organize a team.

8. (Either) the pitcher will throw a ball, or the slugger will hit a home run.
9. (Either) Igor will discover a cure, or the patients will continue to suffer.
10. (Either) Jones will win the battle, or the natives will keep their land.
11. (Either) Brown will defeat the space aliens, or the ship will be swallowed up.
12. (Either) Louise will punish the children, or the parents will decide to let it slide.
13. (Either) the director will fire the cameraman, or the caterers will issue a formal complaint.
14. (Either) the teenagers will leave the party, or the police will tell them to go home.
15. (Either) the acrobat will drop the baton, or the clown will make the children cry.
16. (Either) the waiter will serve the soup, or the cook will have to do it himself.
17. (Either) the distributor will forget to deliver the cookies, or the shoplifters will steal them all.
18. (Either) the investigator will solve the crime, or the victim will continue to wonder what happened.
19. (Either) George will smoke the cigar, or his friend will object to his habit.
20. (Either) the referee will call time out, or the players will get into a scuffle.
21. (Either) the drummer will start the song, or the conductor will tap his baton.
22. (Either) Leo will bring his tools, or the carpenter will make the repairs.

23. (Either) Barbara will organize the files, or the agency will send someone to do it.
24. (Either) Richard will write the paper, or his assistant will do some of the work.

On-line costs for predicting upcoming syntactic structure

Tessa Warren^{a,b,c,d} and Michael Walsh Dickey^{b,d,e}

^a*Learning Research and Development Center, University of Pittsburgh*

^b*Department of Psychology, University of Pittsburgh*

^c*Department of Linguistics, University of Pittsburgh*

^d*Department of Communication Science and Disorders, University of Pittsburgh*

^e*Geriatric Research Education and Clinical Center, VA Pittsburgh Healthcare System*

Abstract

There is considerable psycholinguistic evidence that comprehenders predict the form of upcoming material (Stowe 1986; Frazier & Clifton 1989; Staub & Clifton 2006), and that there is a cost associated with maintaining predictions for mandatory but as-yet-unencountered syntactic elements (Chen et al. 2005). This paper presents a self-paced reading study suggesting that there is a processing cost associated with predicting greater amounts of syntactic structure. In conditions where readers could predict the structure of upcoming material, they were slower to read a word when it spurred the prediction of more structure than when it spurred the prediction of less structure. Two control conditions showed no such difference. These findings indicate that not only is it costly to postulate syntactic structure for incoming material, as has traditionally been assumed (Frazier 1979, 1987; Gorrell 1995; Pritchett 1992), but that it also can be costly to project such material in advance of the bottom-up input.

1. Introduction

An important part of understanding a sentence is the on-line computation of its syntactic structure. This is generally assumed to be done incrementally, more or less word-by-word, as evidenced by almost immediate disruption in response to a perceived syntactic anomaly (e.g., Frazier & Rayner 1982; Fodor et al. 1996). Although much of this computation must rely on bottom-up input, namely the words that make up the sentence, comprehenders also rely on top-down predictions about upcoming structure as they navigate a sentence. For example, encountering an NP that functions as

the subject of a clause will permit the parser to anticipate an upcoming VP. This interaction of top-down predictions and bottom-up input is a characteristic of left-corner parsers (Aho & Ullman 1972), and has led researchers to propose that human comprehenders may use left-corner parsing algorithms (Johnson-Laird 1983; Lewis & Vasishth 2005).

It is commonly assumed that postulating or building syntactic structure is costly for comprehenders. This paper presents evidence from a self-paced reading study suggesting that predicting greater amounts of syntactic structure carries a processing cost. Readers exhibited processing costs (in the form of slower reading times) in conditions where more syntactic structure had to be postulated, at the point in the sentence where the prediction was made. This result is in line with previous findings that comprehenders avoid postulating additional syntactic structure if they can, presumably because postulating such structure is costly (Carlson et al. 2005; Frazier 1979, 1987; Gorrell 1995; Pritchett 1992). The current findings indicate that not only is postulating syntactic structure for incoming material costly, as has traditionally been assumed, but that projecting such material in advance of bottom-up input is costly as well.

1.1 Syntactic prediction

Multiple theories of sentence comprehension include a role for syntactic prediction. The Active Filler Strategy (Frazier & Flores d'Arcais 1989) and its subsequent refinements (Minimal Chain Principle: De Vincenzi 1991, Minimal Dependency Principle: Kazanina et al. 2007) hypothesize that comprehenders anticipate short dependencies between moved elements and their positions of interpretation, or between linguistically dependent elements (such as pronouns) and the elements that license them (their antecedents). These proposals find considerable support from experiments showing that reading is disrupted at positions where a structure could have resolved in such a way as to complete a dependency, but did not and instead left the dependency unresolved (Frazier & Clifton 1989; Stowe 1986; among others). This kind of finding suggests that comprehenders have expectations about upcoming structure and are sensitive to violations of those expectations.

If comprehenders do predict upcoming structure, one would expect such predictions not only to impede comprehension when they are incorrect,

but also to aid comprehension when they are correct. Staub & Clifton (2006) provided psycholinguistic evidence that this is the case. In their experiment, readers read disjunctions that were or were not preceded by the word *either*. Reading times on the second disjunct were faster when the word *either* was present, suggesting that the presence of *either* allowed readers to predict the form of the second disjunct and process it more quickly.

1.2 The cost of prediction

The current experiment builds on Staub & Clifton (2006), but instead of looking for evidence that predicted material is easier to process, it investigates the hypothesis that there is a processing cost associated with projecting predicted syntactic structure, and this cost is related to the amount of structure projected. The idea that projecting syntactic structure is costly is consistent with theories that ascribe a cost to maintaining as yet unfulfilled syntactic predictions across intervening material (e.g., Gibson 1998, 2000). Chen et al. (2005) provided psycholinguistic evidence indicating that maintaining such structural predictions is costly; in their experiments, reading slowed when readers were maintaining more incomplete syntactic dependencies. Chen et al. compared self-paced reading times over the bold-faced regions in (1) and (2). The sentences in (1) differ in how many predictions for verbs must be maintained over the bold-faced region. In (1a), all of the syntactic subjects preceding the bold-faced region are associated with verbs prior to the bold-faced region. However in (1b), the subject *the suspicion* requires a verb that is not encountered until after the bold-faced region, and in (1c) both the suspicion and the knowledge predict verbs that are not encountered until after the bold-faced region. The sentences in (2) vary in whether a prediction for a gap site for the relative pronoun must be maintained across the bold-faced region.

- (1)
 - a. No verbs necessary
The detective suspected that the thief knew that the **guard protected the jewels**.
 - b. One verb needed
The suspicion that the thief knew that the **guard protected the jewels** worried the museum curator.

- c. Two verbs needed
The suspicion that the knowledge that the **guard protected the jewels** came from an insider worried the museum curator.
- (2) a. No wh-gap required
The claim alleging [that the **cop who the mobster attacked** ignored the informant] might have affected the jury.
- b. One wh-gap required
The claim [which the **cop who the mobster attacked** ignored --- the informant] might have affected the jury.

Note that the bold-faced region has the same words and structure across conditions. Nevertheless, reading times were slower across the bold-faced region when more syntactic category predictions were being maintained across it, as demonstrated by slower reading in (1c) than in (1b) than in (1a), and in (2b) than in (2a). This is consistent with the notion that storing syntactic expectations imposes a processing burden, regardless of whether the dependencies are between a subject and verb or a relative pronoun and its interpretation site (see also Gibson et al. 2005; Grodner et al. 2002; Grodner & Gibson 2005).

Given that the storage of syntactic predictions seems to impose a processing cost, the question arises as to whether the establishment of those predictions, or the projection of predicted structure, imposes a processing cost as well. The current experiment addressed this question by building on Staub & Clifton (2006). Specifically, we manipulated the position of *either*, changing the syntactic structure of the first disjunct and thereby varying how much structure a comprehender would posit at the point of predicting the second disjunct. Take the examples in (3a-b).

- (3) a. Jim thinks that either Ally had admired Karen *or* ...
- b. Jim thinks that Ally had admired either Karen *or* ...

The adverb *either* marks the scope of the disjunction headed by *or* (Larson 1985; Schwarz 1999). In (3a), *either* marks the left edge of a clausal disjunction. At the position of *or*, comprehenders can therefore predict an upcoming clause. In (3b), *either* marks the left edge of a simple NP disjunction. At the position of *or* in (3b), comprehenders should generate a prediction for an upcoming NP, a much less complex syntactic structure than in

(3a). Importantly, Staub & Clifton's results indicate that comprehenders do make predictions of exactly these types in structures like these. If it is the case that projecting predicted structure is costly, additional processing costs should be evident on *or* in (3a) as compared to (3b).

2. Study

2.1 Method

2.1.1 Participants

60 native English-speaking undergraduates from the University of Pittsburgh participated in this self-paced reading study. They received partial credit towards an introductory psychology course for their participation. The experiment took approximately 40- 50 minutes to complete.

2.1.2 Materials

The experiment had a 2 x 2 design, crossing adverb type (*either* vs. a modal adverb like *maybe*) with disjunct structure (clause vs. NP). An example is given in (4) below.

- (4) a. Either/clause
Jim thinks that either Ally had admired Karen or Rita had greatly admired Betsy.
- b. Either/NP
Jim thinks that Ally had admired either Karen or Rita but wasn't sure which.
- c. Maybe/clause
Jim thinks that maybe Ally had admired Karen or Rita had greatly admired Betsy.
- d. Maybe/NP
Jim thinks that Ally had admired maybe Karen or Rita but wasn't sure which.

Items began with a name, a clausal-complement verb and *that*. In the clause conditions, *either* or a modal adverb (*maybe*, *perhaps*, *possibly*, or *probably*) followed, preceding a finite clause. In the NP conditions, *either* or the modal adverb appeared within this finite clause, positioned between

the verb and direct object. The subject of this finite clause (*Ally* above) was always a name. Its direct object (*Karen* above) and the noun phrase immediately following the *or* (*Rita* above) were names in some items and definite descriptions in others, but within an item they were always the same type of noun phrase.

The 24 items from this experiment were combined with 120 filler items. Presentation lists were constructed using a Latin square design, so that each list contained one condition from each item and six examples of each condition. The order of presentation within a list was randomized for each participant.

In order to verify that readers predict more syntactic structure in the Either/clause conditions than in the Either/NP conditions, a completion study was run. In this study, items were presented truncated immediately before the *or*, and twelve participants (who did not participate in the self-paced reading study) were asked to complete the sentence fragments with one or more words. Items were counterbalanced across four presentation lists, so each participant saw only one condition of each item. Results are presented in Table 1. There were two important findings. First, participants were considerably more likely to include an *or* in completions for Either conditions (89%) than Maybe conditions (17%). This *or* was almost always the first word in the completion, indicating that *or* is a highly predictable word at this point in the Either conditions, but not in the Maybe conditions. Second, in cases where an *or* was produced, it was considerably more likely to be followed by a clause or verb phrase (74%) than a noun phrase (26%) in the Either/clause condition, but the opposite pattern was apparent in the Either/NP condition, with 86% noun phrases and 16% clauses and verb phrases¹. The few *or* completions in the Maybe conditions were almost exclusively followed by noun phrases. These results confirm that readers strongly predict an *or* in the Either conditions, and predict more structure following that *or* in the Either/clause conditions than in the Either/NP conditions.

¹ Almost all responses were clauses, VPs or NPs. Three responses involved adverbs or adjectives and were coded in the clause and VP category, because they involve more structure than NPs. Three completely ambiguous responses were excluded

	% completions including <i>or</i>	Of <i>or</i> completions, % clause or VP	Of <i>or</i> completions, % NP
Either/clause	81	74	26
Either/NP	97	16	86
Maybe/clause	7	20	80
Maybe/NP	28	5	95

Table 1: Results of completion survey

2.1.3 Procedure

Participants performed word-by-word self-paced reading (Just et al. 1982) in a moving window display on a PC². Sentences were initially presented with dashes replacing all non-space characters. Pressing the space bar revealed each word in turn, and returned the previously visible word to dashes. The time between bar presses was recorded as reading time. A yes-no question tested comprehension after each sentence. Incorrect answers caused the word INCORRECT to briefly appear on the screen.

Participants were given a short practice session at the beginning of the experiment. They were asked to read as naturally as possible and to take incorrect answers as an indication to read more carefully.

2.2 Results

Comprehension was high during the experiment; on average, participants correctly answered approximately 85% of comprehension questions across the entire session. Reading times that were longer than 1500 or shorter than 160 milliseconds were excluded from analyses to eliminate outliers. This procedure eliminated approximately two percent of the data. Figure 1 presents average reading times for each condition, with some words grouped into regions for the purposes of display.

The critical word in this experiment was the *or*, because it is the earliest point at which a comprehender might build a syntactic projection for the second disjunct. Average reading times for this word are presented in Figure 2³. The data were subjected to repeated-measures ANOVAs using

² The software was E-Prime (Schneider et al. 2002)

³ One condition of one item had an error in the placement of either and was therefore treated as

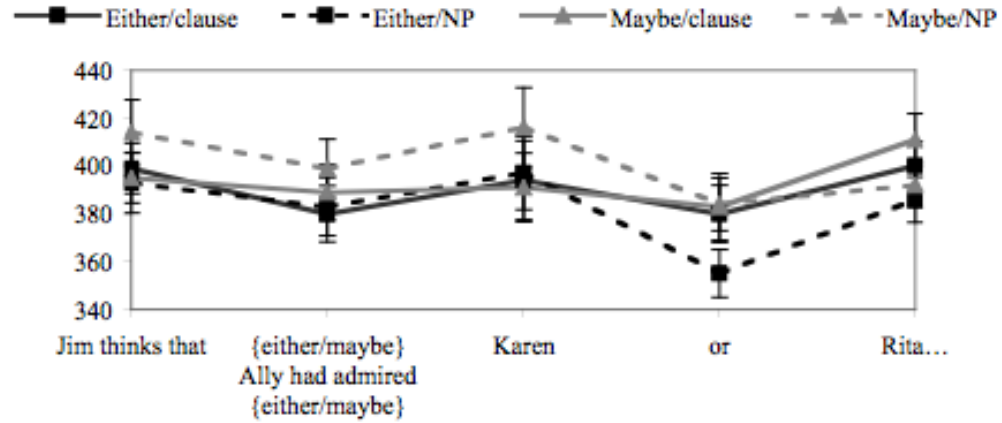


Figure 1: Average reading times per word in milliseconds

both participants (F1) and items (F2) as random factors. 2x2 ANOVAs on reading times at the critical *or* revealed no main effects, but an interaction between the presence of *either* and disjunct structure that was fully reliable in the analysis by participants, although not in the one by items ($F(1,59)=4.5$; $p=.039$; $F_2 < 1$). Pair-wise tests comparing just the two *Either* conditions indicated they were reliably different in the participants analysis and marginally reliable in the items analysis, with the *Either/NP* condition faster than the *Either/clause* condition ($F(1,59)=6.8$; $p=.012$; $F_2(1,22)=3.45$; $p=.08$). The two *Maybe* conditions did not differ. ANOVAs showed no reliable reading time differences between conditions on either the word preceding or the word following the critical *or* (all $ps > .1$).

2.3 Discussion

The results of this experiment support the hypothesis that comprehenders predict upcoming syntactic structure on-line during reading, and that the cost of doing so is related to the amount of structure being predicted. Following *either*, reading times were longer on *or* when the upcoming disjunct was anticipated to be a clause than when it was anticipated to be a NP. Importantly, in two corresponding conditions where the *either* was replaced by a modal adverb that established neither a strong prediction for a disjunction

missing data and excluded in analyses by items.

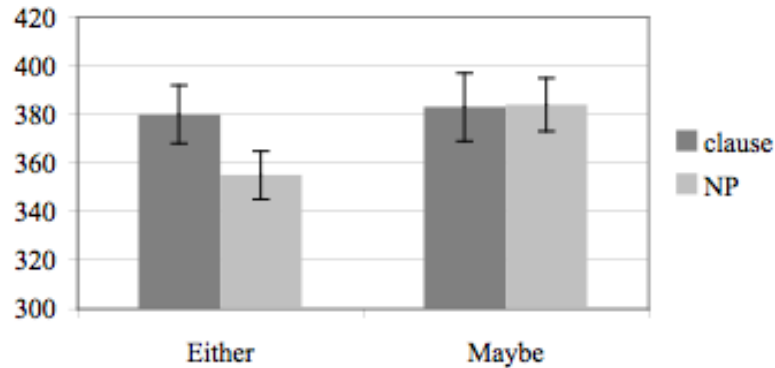


Figure 2: Average reading times on “or” in milliseconds.

nor parallelism between the disjuncts (viz. Yoshida & Dickey 2008), reading times on *or* were not different. This data pattern suggests that readers read more slowly at the point of establishing an expectation for an upcoming clause than when establishing an expectation for a simple NP. In cases where they had no expectation as to the form of the second disjunct, there was no modulation of reading times on *or*.

One potential concern about the reading time pattern is that reading times on the *or* were similar in the Either/clause and Maybe conditions. If syntactic prediction were the only factor influencing reading times at this word, the Maybe conditions should have patterned with the Either/NP conditions, as little syntactic material is predicted in any of these conditions. However, the completion study indicated a second factor likely influencing reading times on this word, namely the relative predictability of *or*. Given that *or* was highly predicted in the Either conditions but not in the Maybe conditions, and that highly predicted words are read more quickly than less predicted words (Rayner & Well 1996), the observed pattern may simply reflect the combination of these two factors. Specifically, the prediction of more syntactic material may have increased reading times on the *or* in the Either/clause condition as compared to the other three conditions, but the higher predictability of *or* in the Either conditions may also have depressed reading times relative to the Maybe conditions, causing the observed pattern. It is worth noting that this account does not explain the

observed reading-time difference between the Either/NP and Either/Clause conditions. The presence of *or* is equally predictable in both these conditions, so the observed difference must be due to other factors, in particular the prediction of upcoming structure which is licensed in the presence of *either*.

The design of the current experiment eliminates some potential alternative explanations for the data patterns. If long reading times on the *or* in the Either/clause condition as compared to the Either/NP condition had been due to wrap-up processing associated with the completion of a clause (e.g., Hirotsani et al. 2006; Warren et al. 2009), the same effect should have been evident in the Maybe conditions. No such effect was apparent. One possibility that the design does not rule out, but that seems unlikely given the data patterns and current understanding of the relationship between *either* and *or*, is that the difference between the two *either* clauses might reflect the difficulty of establishing a dependency between the *either* and *or*. There is considerable evidence that reading slows at positions where readers must resolve long-distance dependencies (e.g., Gibson 1998; Grodner & Gibson 2005). If, at the point of reading *or*, readers must establish a dependency between the *either* and *or*, this dependency would be longer in the Either/clause condition than the Either/NP condition. This account could therefore predict the difference between the Either conditions. There is no reason to expect that there would be a similar dependency between the modal adverb and *or*, so this account would also correctly predict a lack of difference between the two Maybe conditions. However, as noted above, previous work has shown that holding a syntactic prediction over intervening material generates a reading-time penalty. Readers are slower to read intervening material when the parser must hold a syntactic prediction in memory (Chen et al. 2005; Wanner & Maratsos 1978). If there were a direct syntactic dependency between *either* and *or*, conditions with *either* should have elicited longer reading times on the words preceding *or* than conditions with a modal adverb like *maybe*. This was not the case in the current data; in fact, average reading time per word across the first disjunct was longer in the Maybe conditions than in the Either conditions.

There are also linguistic arguments against assuming that there is a direct linguistic dependency relation between *either* and *or*. For example, some have argued that *either* might be moved from its initial position adja-

cent to *or* to its surface position marking the scope of the disjunction, either through an operation similar to wh-movement (Larson 1985) or one similar to Quantifier Raising (Munn 1993). However, there is reason to doubt this analysis of the relationship between *either* and *or*, on both linguistic and psycholinguistic grounds.

Linguistically, the direct-dependency approach to *either-or* incorrectly predicts that the *either-or* relation should obey the same locality constraints as wh-movement or QR. As Schwarz (1999) shows, it does not. A finite clause selected by a clausal-complement verb is not a barrier to wh-movement, but it is a barrier to the *either-or* relationship (5):

- (5) a. *What* did John say that Mary was cooking or eating __ last night?
 b. ?? John said *either* that Mary was cooking *or* eating beans last night.

Even claiming that only infinitival clauses permit *either* movement, as Larson (1985) does to account for cases like (5), is incorrect. The infinitival clause introduced by *for* in (6a-b) is not a barrier to wh-movement, but it is a barrier to the relationship between *either* and *or*:

- (6) a. *What* did John want for Mary to cook or eat __ last night?
 b. ?? John *either* wanted for Mary to cook *or* eat beans last night.

Similarly, the *either-or* relationship is blocked in cases where QR is not. The wide-scope reading of the quantified noun phrase *every guest* is available in (7), indicating that it can undergo quantifier raising to a position to the left of the indefinite *something*.

- (7) Something pissed *every guest* off.

However, *either* cannot be related to *or* when it occurs to the left of something in the parallel examples in (8a-b):

- (8) a. Something pissed *either* Bill *or* Sue off.
 b. ?? *Either* something pissed Bill *or* Sue off.

Examples like these lead Schwarz (1999) to conclude that there cannot be a direct linguistic dependency between *either* and *or*. Instead, *either*

is an optional adverbial element, one which marks the left edge of the disjunction when it is present (Johannessen 2005; Neijt 1979; Sag et al. 1985).⁴

Psycholinguistically, the direct-dependency view of *either-or* incorrectly predicts that there should be a reading-time disadvantage for conditions with *either* before encountering *or*. Both this fact and the linguistic facts above indicate that *either* and *or* are not linked in a direct grammatical dependency. Such a dependency is therefore unlikely to be the source of the reading time difference on the *or* in the two Either conditions.

3. Conclusion

Comprehenders must assign a syntactic structure to incoming words in order to interpret a sentence, and there is considerable evidence that they do so incrementally. The current study provides additional evidence that comprehenders generate syntactic structure not only incrementally but anticipatorily (Yoshida & Dickey 2008), and that there are costs associated with projecting such structure. Participants were slower to read a sentence at the point where they could generate a prediction for more upcoming syntactic structure. This result is consistent with other findings indicating that comprehenders actively anticipate the structure of upcoming material (Stowe 1986; Frazier & Clifton 1989; Staub & Clifton 2006) and that there are costs associated with maintaining unsatisfied syntactic predictions (Chen et al., 2005). The current results provide direct evidence that is consistent with the long-standing assumption that comprehenders prefer simpler structures precisely because such analyses allow them to avoid costly structure-building operations (viz. Frazier 1979).

References

- Aho, Alfred V. & Jeffrey D. Ullman (1972). *The theory of parsing, translation and compiling*. Prentice-Hall Inc., Englewood Cliffs, N.J.
- Carlson, Katy, Michael Walsh Dickey & Christopher Kennedy (2005).

⁴ Johannessen (2005) claims that the optional *either* acts as a focus particle when it is present, licensed by a correlative functional project CorP. Her account also postulates a combination of overt and covert movement of *either*, which would seem vulnerable to the same sorts of criticisms Schwarz (1999) levels against related movement analyses (Larson 1985; Munn 1993).

- Structural economy in the processing and representation of gapping sentences. *Syntax* 8:3, pp. 208–228.
- Chen, Evan, Edward Gibson & Florian Wolf (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language* 52, pp. 144–169.
- De Vincenzi, Marica (1991). *Syntactic Parsing Strategies in Italian*. Kluwer, Dordrecht, The Netherlands.
- Fodor, Janet Dean, Weijia Ni, Stephen Crain & Donald Shankweiler (1996). Tasks and timing in the perception of linguistic anomaly. *Journal of Psycholinguistic Research* 251, pp. 25–57.
- Frazier, Lyn (1979). *On Comprehending Sentences: Syntactic Parsing Strategies*. Indiana University Linguistics Club, Bloomington, IN.
- Frazier, Lyn (1987). Sentence processing: a tutorial review. Coltheart, M. (ed.), *Attention and Performance XII*, Erlbaum, Hillsdale, NJ, pp. 559–585.
- Frazier, Lyn & Charles Clifton, Jr. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4:2, pp. 93–126.
- Frazier, Lyn & Giovanni Flores d'Arcais (1989). Filler driven parsing: a study of gap filling in dutch. *Journal of Memory and Language* 28, pp. 331–344.
- Frazier, Lyn & Keith Rayner (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14, pp. 178–210.
- Gibson, Edward (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68, pp. 1–76.
- Gibson, Edward (2000). Linguistic complexity: Locality of syntactic dependencies. Miyashita, Y., A. Marantz & W. O'Neil (eds.), *Image, language, brain*, MIT Press, Cambridge, MA, pp. 95–126.
- Gibson, Edward, Timothy Desmet, Daniel Grodner, Duane Watson & Kara Ko (2005). Reading relative clauses in English. *Cognitive Linguistics* 16:2, pp. 261–290.
- Gorrell, Paul (1995). *Syntax and Parsing*. Cambridge University Press, Cambridge, UK.
- Grodner, Daniel & Edward Gibson (2005). Consequences of the serial na-

- ture of linguistic input for sentential complexity. *Cognitive Science* 29, pp. 261–290.
- Grodner, Daniel, Edward Gibson & Susanne Tunstall (2002). Syntactic complexity in ambiguity resolution. *Journal of Memory and Language* 46, pp. 267–295.
- Hirotsani, Masako, Lyn Frazier & Keith Rayner (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language* 54, pp. 425–443.
- Johannessen, Janne Bondi (2005). The syntax of correlative adverbs. *Lingua* 115, pp. 228–238.
- Johnson-Laird, Phillip N. (1983). *Mental Models*. Cambridge University Press, Cambridge, UK.
- Just, Marcel Adam, Patricia A. Carpenter & Jaqueline D. Woolley (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 11:2, pp. 228–238.
- Kazanina, Nina, Ellen F. Lau, Moti Lieberman, Masaya Yoshida & Colin Phillips (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language* 56, pp. 384–409.
- Larson, Richard K. (1985). On the syntax of disjunction scope. *Natural Language and Linguistic Theory* 3, pp. 217–264.
- Lewis, Richard L. & Shravan Vasishth (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29, pp. 375–419.
- Munn, Alan B. (1993). *Topics in the syntax and semantics of coordinate structures*. Ph.D. thesis, University of Maryland, College Park.
- Neijt, Anna (1979). *Gapping: A Contribution to Sentence Grammar*. Foris, Dordrecht, The Netherlands.
- Pritchett, Bradley L. (1992). *Grammatical competence and parsing performance*. University of Chicago Press, Chicago.
- Rayner, Keith & Arnold D. Well (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review* 3, pp. 505–509.
- Sag, Ivan, Gerald Gazdar, Thomas Wasow & Steven Weisler (1985). Coordination and how to distinguish categories. *Natural Language and*

Linguistic Theory 3, pp. 117–171.

- Schneider, Walter, Amy Eschman & Anthony Zuccolotto (2002). *E-Prime user's guide*. Psychology Software Tools, Inc.
- Schwarz, Bernhard (1999). On the syntax of either . . . or. *Natural Language and Linguistic Theory* 17, pp. 339–370.
- Staub, Adrian & Charles E. Clifton, Jr. (2006). Syntactic prediction in language comprehension: Evidence from either. . . or. *Journal of Experimental Psychology: Learning, Memory and Cognition* 322, pp. 425–436.
- Stowe, Laurie (1986). Evidence for on-line gap location. *Language and Cognitive Processes* 1, pp. 227–245.
- Wanner, Eric & Michael Maratsos (1978). An ATN approach in comprehension. Halle, Morris, Joan Bresnan & George A. Miller (eds.), *Linguistic theory and psychological reality*, MIT Press, pp. 119–161.
- Warren, Tessa, Sarah J. White & Erik D. Reichle (2009). Investigating the causes of wrap-up effects: Evidence from eye movements and e-z reader. *Cognition* 111:1, pp. 132–137.
- Yoshida, Masaya & Michael Walsh Dickey (2008). Prediction in parsing: Either syntactic or semantic. Poster presented at 21st annual CUNY conference on Sentence Processing, Chapel Hill N.C.

