The Role of Context in the Resolution of
Quantifier Scope Ambiguities

ELISABETH VILLALTA
Utrecht Institute of Linguistics—OTS

Abstract
This paper presents experimental results that elucidate some aspects of semantic processing, i.e. of the system that allows perceivers to associate an interpretation to a sentence on-line. The phenomenon under investigation is the resolution of quantifier scope ambiguities. Sentences containing multiple quantifiers (i.e. everybody, some musician, etc.) are known to give rise to several interpretations. The question addressed in this work is how this kind of ambiguity is resolved in the on-line process of constructing an interpretation for a sentence. The research reported here concentrates exclusively on English and French interrogative sentences, and in particular on the case of ambiguous how many questions that contain a universally quantified subject, every N.

The central results of this paper are the following. First, quantifier scope preferences are shown to be problematic for the most straightforward extension of an economy-based model to the processing of meaning, as evidenced by questionnaire studies in English and in French. Second, a model is elaborated in which the attested scope preferences are determined by the interaction with context. The results from a self-paced reading study in English indicate that context plays a crucial role in the processing of scope ambiguity. Third, while incremental context interactive models have been claimed to induce immediate resolution of structural ambiguity (Crain & Steedman 1985; Altmann & Steedman 1988; and others), it is argued here that the interaction with context can also delay such ambiguity resolution, as evidenced by the results of the English self-paced reading study. Finally, comparison of the two languages, English and French, sheds further light onto the phenomenon under investigation.

1 INTRODUCTION

Considerable research in psycholinguistics has addressed the question of how speakers/listeners construct syntactic representations for sentences on-line. The question of how these syntactic representations are interpreted on-line, however, has not yet received the same attention. This paper aims at elucidating some aspects of the mechanism that
disambiguates syntactic representations and provides the representations that are interpreted by the semantic component of the parser. In particular, the area under investigation is the process of resolving quantifier scope ambiguity.

Sentences containing multiple quantifiers (elements of the kind *everybody, some musician, many pieces* etc.) are known to give rise to several interpretations. The specific question addressed in this work is how this kind of ambiguity is resolved in the on-line process of constructing an interpretation for a sentence. I concentrate exclusively on interrogative sentences, and in particular on the case of ambiguous *how many* questions that contain a universally quantified subject, *every N*. An example is given in (1).

(1) How many pieces did every musician play?

Psycholinguistic work on the resolution of quantifier scope ambiguity has so far mainly focused on declarative sentences (e.g. Kurtzman & MacDonald 1993; Tunstall 1997). In questions, resolution of quantifier scope ambiguity arises when the interrogative constituent (i.e. *how many pieces*) interacts with another quantifier in the sentence (i.e. *every musician*). For example, the question illustrated in (1) can receive several interpretations (and thus receive different possible answers), depending on whether the quantifier encoded in the interrogative constituent *n-many pieces* is interpreted in the scope of the quantifier *every musician* or not. A more detailed presentation of the relevant ambiguity will be given shortly, in Section 2.

I will follow the view adopted in generative grammar where it is assumed that quantifier scope ambiguity is a structural ambiguity. According to standard assumptions, this ambiguity is due to the fact that a sequence of quantifiers is not necessarily interpreted with its surface order. In generative grammar, a sentence with several quantifiers receives several syntactic representations, so-called logical forms (LFs), where the order of the quantifiers varies accordingly. These LFs are syntactic representations that are derived from the surface representation of a sentence with the help of covert syntactic movement operations. Since May (1985), two commonly used covert movement operations are Quantifier Raising (QR) and Quantifier Lowering (QL). As a consequence of QR and QL, in each LF representation, the quantifiers appear in the structural configuration required by the corresponding interpretation. LF is thus a level of syntactic representation where scope is overtly disambiguated. This level of representation mediates between surface syntactic structures.
and meaning, in other words it provides the input to semantic interpretation.¹

Under this view, a sentence processing model has to contain a mechanism that constructs the LF representations of an incoming surface representation and chooses the LF that yields the preferred interpretation of the sentence. The research presented in this paper offers experimental results that elucidate some aspects of this mechanism.

The central results of this paper are the following. First, quantifier scope preferences in questions will be shown to be problematic for the most straightforward extension of a processing model based on Economy assumptions, namely a model in which quantifiers are integrated into the LF representations in the order in which they appear in surface structure. The results from questionnaire studies in English and French indicate that such a model does not make the correct predictions for how many questions. This, of course, raises a number of questions on how comprehenders associate an interpretation on-line to the sentences they perceive.

Second, the results from a self-paced reading study in English provide on-line evidence for a model in which scope ambiguity is resolved through the interaction with context. This is an important finding, since, in the psycholinguistic literature, it has not yet been demonstrated that the resolution of quantifier scope ambiguity is dependent on the context that precedes the sentence. The few psycholinguistic studies that have dealt with the resolution of quantifier scope ambiguity (cf. Kurtzman & MacDonald 1993; Tunstall 1997) have used sentences in isolation and thus have not examined the role of the preceding discourse context. Given the results of the present work, the elaboration of a more generalized theory of quantifier scope resolution will require that preceding context be taken into consideration and carefully controlled in experimental work.

Finally, the results from the self-paced reading study furthermore indicate that interaction with context can delay ambiguity resolution. This result contrasts with the important well known claim that incremental context-interactive models induce immediate resolution of

¹ May (1985) proposes that LF is ‘that level of representation which interfaces the theories of linguistic form and interpretation. On this view, it represents whatever properties of syntactic form are relevant to semantic interpretation—those aspects of semantic structure that are expressed syntactically.’ (May 1985: p. 2). The assumption that LF is the level of representation that feeds semantic interpretation has been widely adopted in recent semantic literature. In particular, this view is proposed in Heim & Kratzer (1998) and Chierchia & McConnell-Ginet (1990). It is also presented and discussed in de Swart (1998: p. 100–103) and in Partee et al. (1993: p. 332). I will adopt this view here, even though it is of course not the only possible view of how syntax and semantics interact.
structural ambiguity (Crain & Steedman 1985; Altmann & Steedman 1988; and others).

The empirical basis of this paper consists of data from English and French how many questions. Along with the usual how many construction, French makes use of a split construction in which only combien (‘how many’) is fronted. The case of French is particularly informative, since the split and non-split how many constructions differ in their possible scope configurations. While the non-split construction does give rise to scope ambiguity, the split construction does not. Comparison of the two languages will thus help to shed more light on the processes involved in the resolution of quantifier scope ambiguity.

The structure of this paper is as follows. Section 2 presents the phenomenon of scope ambiguity in how many questions. Section 3 lays out the basic assumptions of the parsing model that are adopted. In Section 4, a first hypothesis for the scope resolution in how many questions is proposed. This hypothesis is based on the assumption that the parser obeys certain economy principles when constructing an LF representation of a sentence. The section concludes with a questionnaire study in English that provides strong evidence against this hypothesis. In Section 5, the French data are presented. The results from a French questionnaire study further strengthen the conclusion from Section 4. In Section 6, the central proposal is developed. It is argued that scope resolution is determined by the interaction with context. Furthermore, evidence is provided which suggests that interaction with context does not necessarily accelerate ambiguity resolution. The results of a self-paced reading study in English indicate that, in certain cases, the interaction with context can also result in the delay of such ambiguity resolution.

2 THE PHENOMENON: QUANTIFIER SCOPE AMBIGUITIES IN HOW MANY QUESTIONS

In this section, the empirical facts about quantifier scope ambiguity in how many questions are presented. Two possible readings of these questions are discussed, and their two underlying LF representations are formulated.

A particular interpretation of a question is revealed by the answer it requires in a given context. How many questions that contain a universally quantified subject can receive at least two interpretations. Imagine the scenario described in (2).

(2) In the music department, three trumpet students had to pass an exam last week. Every student had to play six pieces. The only
requirement they had was that among these there were two pieces that everybody had to play: 'Round Midnight' and 'The days of Wine and Roses'. For the rest, the students were free to choose what they preferred.

In the context of the situation described above, the question in (3) can be answered in two possible ways, namely, as in (3a) and (3b).²

(3) How many pieces did every student have to play at the exam?
   Possible answers  a. Six pieces
                   b. Two pieces

The example presented in (3) is a case of scope interaction between \textit{n}-many pieces and the subject quantifier \textit{every student}. The fact that \textit{how many} questions are potentially ambiguous in scope was noted first in Longobardi (1987) and Kroch (1989) and later explored by many others (Cinque 1990; de Swart 1992; Szabolcsi & Zwarts 1993; Dobrovie-Sorin 1993; Rullmann 1995; Cresti 1995; Beck 1996; Honcoop 1998; and others). With a predicate logic, the two possible interpretations can be represented as in (4a) and (4b). These two representations differ only in the order of the two quantifiers (where the operator card(Y) returns the cardinality of the set Y).³

(4) a. For which number \( n \): \( \forall x \text{ student}(x) \Rightarrow \exists Y \text{ pieces}(Y) \land \text{card}(Y)=n \land \text{played}(Y)(x) \)
   Answer: Six pieces
   b. For which number \( n \): \( \exists Y \text{ pieces}(Y) \land \text{card}(Y)=n \land \forall x \text{ student}(x) \Rightarrow \text{played}(Y)(x) \)
   Answer: Two pieces

The LF representations that I will propose for these two interpretations of a \textit{how many} question are given in (5a) and (5b).

² Another possible answer in this context is a so-called \textit{cumulative} answer in which the total number of pieces is counted (in this example eighteen). I will not discuss this interpretation here. A further possibility is the so called \textit{pair-list} answer. This kind of answer provides a list with information about each individual (e.g. \textit{Jason played six pieces, Joyce played three pieces and Brian played ten pieces}). In the context provided in (2), a \textit{pair-list} answer is however infelicitous, since everybody played the same amount of pieces and no information about the individuals was given.

³ Note that the two representations in (4a) and (4b) are in fact not logically independent. That is, with this formalization, in a given situation, every possible answer to (4b) is also a possible answer to (4a). This is so since, under this formalization, 'six pieces', 'five pieces', 'four pieces', 'three pieces', 'two pieces' and 'one piece' are correct answers to the question in (4a). In order to maintain two distinct representations for the two readings, we need to prevent the 'two pieces' answer to be an appropriate answer to the representation of the question in (4a). To achieve this, I propose to adopt the Gricean principle of Informativeness. This principle requires that a question be answered with a maximally informative answer. Under this principle, only 'six pieces' is an appropriate answer to (4a), not 'two pieces'.
A few comments are in order to explain why these two LF representations are the adequate underlying syntactic representations that lead to the corresponding interpretations in (4a) and (4b).

For concreteness, I assume a semantics for questions along the lines of Hamblin (1971) and Karttunen (1977), where the denotation of a question is the set of propositions which constitute possible answers to that question. For instance, if it is the case that Jason, Joyce and Brian came to the rehearsal, the sentence *Who came to the rehearsal?* denotes the set \{‘Jason came to the rehearsal’, ‘Joyce came to the rehearsal’, ‘Brian came to the rehearsal’\}. Hamblin and Karttunen adopt an approach in which the *wh*-phrase contains an existential quantifier (‘*For which x*’), and a question operator *Q*, situated in an appropriate position in the structure, which turns the proposition ‘x came to the rehearsal’ into a set of propositions. In other words, the question *Who came to the rehearsal?* denotes a set of propositions that are identical to ‘x came to the rehearsal’ for some person x.

It has been noticed in the literature that *how many* phrases are semantically more complex than other *wh*-phrases in that they involve two independent scope bearing elements (cf. Cresti 1995; Rullmann 1995; Beck 1996, among others). A question of the kind *How many pieces did Jason play?* is interpreted as the set of propositions ‘Jason played n pieces’ for some number n. Appropriate representations for the possible interpretations can only be determined if *how many* phrases are decomposed into two quantificational elements, one encoding an existential quantifier over numbers n (‘*For which n*’) and a second encoding an existential quantifier over sets. The semantically interrogative part ‘*For which n*’ has to be separated from the existential quantifier over sets, since the latter is required to be interpreted inside the scope of the question operator that turns the proposition into a set of propositions. If the existential quantifier were to be interpreted, like the quantifier over numbers, outside the scope of the question operator, the resulting meaning of the question *How many pieces did Jason play?* would be a set of propositions of the kind ‘ Jason played X’, where X are pieces and the set X has some cardinality n, for instance {‘Jason played *Round Midnight*, *The days of Wine and Roses*

(5) a. LF1: [CP How_{n} [C Q[s every student, [[t_{n}-many pieces], [v_{p} t_{i} played t_{j}]]]]]
   LF1 answer: Six pieces

b. LF2: [CP How_{n} [C Q[[t_{n}-many pieces], [IP every student, [v_{p} t_{i} played t_{j}]]]]]
   LF2 answer: Two pieces
and So What’, ‘Jason played The days of Wine and Roses and So What, etc.). This clearly does not correspond to the appropriate set of possible answers to that question. Since the question operator is assumed to be interpreted in the Complementizer position C0, the existential quantifier over sets necessarily has to be interpreted below that position, while the quantifier over numbers takes matrix scope. For convenience, I repeat the two LF representations here in (6a) and (6b). These representations only differ in the order of their quantifiers (in these representations, Q stands for the abstract question operator situated in the Complementizer-position, and t stands for the trace of an element that has been moved out of its surface position; coindexing indicates which element the trace is associated to).

(6)  
\[ \text{a. } \text{LF1: } [\text{CP } \text{How}_n [C Q[\text{every student}] [t_n-many pieces] [\text{VP}_t \text{played } t_j]]]]\]  
LF1 answer: Six pieces  
\[ \text{b. } \text{LF2: } [\text{CP } \text{How}_n [C Q[t_n-many pieces] [\text{IP}_t \text{every student}] [\text{VP}_t \text{played } t_j]]]]\]  
LF2 answer: Two pieces

Following Cresti (1995), in the possible LF representations of a how many question, the quantifier over numbers is encoded in how and takes matrix scope as required by the question meaning, while the quantifier over sets is encoded in n-many N and lowered below the question operator in C0 (by Quantifier Lowering). n-many N can be lowered into two possible landing sites which results in the two Logical Forms stated in (6a) and (6b).

In what follows, I will limit myself to the use of the corresponding syntactic representations, the Logical Forms (6a) and (6b) that lead to the interpretations represented in (4a) and (4b). I will not provide a full semantics of how many questions, since the corresponding semantic representations can be readily computed from the LF representations. The technical details are irrelevant for the purposes of this paper. I refer the reader in particular to Cresti (1995) and references therein to a discussion of a semantic derivation for how many questions.

Given these two LF representations, the following question arises. During sentence processing, how does the parsing mechanism determine the LF representation that is to be associated to the sentence (reflecting the preferred interpretation of the question)? Before formulating a hypothesis concerning the resolution of scope ambiguity in how many questions, I turn to the underlying assumptions of the processing model adopted in this paper.
3 THE ASSUMPTIONS OF THE PROCESSING MODEL

In this section, I lay out the basic assumptions for the parsing model that I adopt. In order to compute an interpretation resulting from a particular scope configuration, it is necessary that the parser build the corresponding LF representation. This implies that, when the quantificational elements are not in the appropriate configuration at surface structure, the parser has to build the structural representation that can feed semantic interpretation. The LF representation is computed along with the syntactic surface structure as the incoming words are perceived on-line.

In previous literature, it has been argued that scope ambiguity resolution requires a model of parallel evaluation. Kurtzman & MacDonald (1993) propose a model for the resolution of scope ambiguity in declarative sentences, where the different possible interpretations are initially considered in parallel. A set of scope principles then determines which representation (corresponding to the preferred interpretation) is ultimately selected. Crain & Steedman (1985) and Altmann & Steedman (1988), similarly, propose that the resolution of structural ambiguity in a context interactive model requires parallel evaluation of partial interpretations. They argue that the appropriate interpretation of an ambiguous sentence in a particular context can only be chosen through comparison of alternatives. In their view, a single interpretation cannot be rejected on grounds of implausibility, but only in comparison with some more plausible alternative. They therefore claim 'that weakly interactive processors must by definition propose syntactic alternatives for semantic and pragmatic adjudication in parallel.' (Altmann & Steedman 1988: p. 208).

The main concern of these models is the mechanism that chooses among the possible interpretations of a sentence. How the actual disambiguated representations are constructed is however not made explicit. The model that I propose here focuses on the mechanism that actually constructs the disambiguated representations. Once such a mechanism is made explicit, it will be shown that, in the case of ambiguity resolution in how many questions, parallel evaluation of several partial interpretations becomes unnecessary. Specifically, I will argue that, in an on-line context interactive model, at the point of ambiguity, the context determines the construction of a single LF representation, without the need to compare different alternatives.

4 THE MINIMAL COST HYPOTHESIS

In this section, I formulate a first hypothesis for the computation of LF representations of an ambiguous how many question.
A vast body of literature has argued for a parser that obeys economy principles when constructing surface representations (cf. the Minimal Attachment Principle, Frazier 1978, Simplicity, Gorrell 1995, the Minimal Chain Principle, de Vincenzi 1991, and many others). In this work, I adopt the hypothesis that the construction of LF representations is also governed by economy principles. Indeed, the few recent studies that have been explicit about how corresponding LF representations are associated to surface representations, argue that the parser first chooses to construct the LF that requires minimal changes from the surface representation (e.g. the Principle of Scope Interpretation by Tunstall 1997 and the Minimal Lowering Principle by Frazier 1999). Of particular interest is the proposal made by Tunstall (1997), which explicitly addresses the question of how LF representations for sentences with multiple quantifiers are constructed. Her Principle of Scope Interpretation states that the preferred interpretation of a sentence corresponds to the LF that differs minimally from the surface structure.

In a similar vein, I examine the hypothesis that the parser first chooses to construct the LF that has minimal cost. This notion can, of course, be defined in many different ways. Here, I investigate the case in which minimal cost mirrors Tunstall’s concept of minimal changes from the surface representation. To make such a proposal explicit, I define the following cost function: the cost of an LF corresponds to the number of permutations that are necessary to derive the order of the quantifiers from their surface order. Thus, an LF in which the quantifiers have been permuted with respect to surface order has higher cost than an LF where the order of such elements is preserved (e.g. \( \text{cost}(Q1Q2) = 0 < \text{cost}(Q2Q1) = 1 \)). For clarification, I repeat the two LF representations in (7a) and (7b). In LF1, the quantifiers are reversed with respect to their surface order; in LF2, their order is preserved.

\[
\begin{align*}
(7) & \quad \text{a. LF1: } [\text{CP Hown } [C Q [IP every studenti [[t_n-many piecesj] \text{VP}_{ti} \text{played } t_j]]]]] \\
& \quad \text{LF1 answer: ‘Six pieces’}
\end{align*}
\]

\[
\begin{align*}
(7) & \quad \text{b. LF2: } [\text{CP Hown } [C Q [IP [t_n-many piecesj] [Ipevery studenti \text{VP}_{ti} \text{played } t_j]]] ] \\
& \quad \text{LF2 answer: ‘Two pieces’}
\end{align*}
\]

According to such a cost function, we expect the parser to construct the LF representation that respects the order of the quantificational elements in which they appear at Surface Structure. This leads to the formulation of the Minimal Cost Hypothesis as stated in (8).
Minimal Cost Hypothesis
When processing a how many question, the parser first computes LF2, because it has less cost than LF1 (cost(LF2) < cost(LF1)).

In order to test the Minimal Cost Hypothesis, a questionnaire study was designed. The interpretation that perceivers choose for an ambiguous question can be determined rather straightforwardly. It is sufficient to ask them to answer the question after having read a story. In the questionnaire study, test-questions were thus preceded by short stories. These were set up in a way that both interpretations of the question were possible answers. This strategy not only determines directly which interpretation participants associate to the question, but has the further advantage that the context in which the question is processed can be controlled.

4.1 Experiment 1
This experiment consisted of a questionnaire study that tested the prediction of the Minimal Cost Hypothesis laid out in (8). That is, it examined whether native English speakers do in fact prefer to associate an LF2 representation to a how many question. The questionnaires presented how many questions preceded by a short story that permitted both interpretations of the question. Each story described a scenario in which several individuals engaged in an activity. This activity involved manipulating sets of objects or entities, some of which were the same for all individuals mentioned. An example is given in (9).

(9) Three friends went to the last Film Festival in Montreal. Altogether, each of them saw ten movies. When comparing what they had seen at the end, they realized that there were four movies that they all had seen.

Question: How many movies did everybody see at the last Film Festival in Montreal?

The story in (9) presents information about the number of movies that each of the friends saw (LF1 answer) as well as the number of particular movies that were seen by everybody (LF2 answer). These two pieces of information were mentioned explicitly in all stories and we expected participants to choose one of the two corresponding answers. The stories were meant to be constructed as ‘neutral’ as possible by not presenting either of the two pieces of information as more salient or relevant; the verb was kept the same in the question and in both
sentences that contained the relevant information. In order to make the information in the stories as unambiguous as possible, statements of the kind everybody saw n movies were avoided, and replaced, when possible, by less ambiguous statements (e.g. each of them saw n movies for the LF1 information, there were n movies that were seen by everybody for the LF2 information). This would ensure that we could correctly interpret participants’ answers. Since the frames corresponding to LF1 and LF2 information were different from the frame used in the question, no bias was expected. Only in one story were the frames introducing LF1 and LF2 information identical to the frame in the question. But, since here all three frames were the same, no bias was expected either.

Two factors were manipulated: the order in which the cardinality information was presented in the text, and the non-partitive/partitive nature of the wh-element (how many N versus how many of the N).

The first factor was introduced to eliminate the possibility that participants would simply answer with the last information they recalled from the text. In the counterpart stories, the order of the cardinality information was presented in the reverse order, as in (10).

(10) Three friends went to the last Film Festival in Montreal. At the end, when comparing what they had seen, they realized that there were four movies that they all had seen. Altogether, each of them saw ten movies.

The second factor (partitive/non-partitive wh-phrases) was introduced for generality reasons. There is a possibility that the partitive/non-partitive nature of the how many phrase plays a determining role for its scope preferences. It is well known that non-interrogative partitive Noun Phrases do not have identical scope preferences to the corresponding non-partitive Noun Phrases. Initial intuitions, indeed, seemed to suggest that partitive how many of the N would disfavour more strongly an LF1 answer. If there was a difference, the Minimal Cost Hypothesis would not be able to capture it and would need reformulation. Comparison of partitive and non-partitive phrases would thus provide more relevant information about the parsing mechanism, and elucidate whether there is more than one principle at play in the construction of an LF representation.

The filler items were stories followed by which questions (e.g. Which pieces did every student play?). In these stories, instead of the information on the cardinality of the sets, the names of the manipulated objects/entities were provided. Ambiguous which questions can be answered either with a pair-list answer (e.g. Jason played Round Midnight and Autumn Leaves, Joyce played Round Midnight and Summertime, Rebecca
played *Round Midnight* and *The Rainbow People*), or with a single answer (e.g. ‘*Round Midnight*’). These stories served as filler items, but they also permitted comparison between how many questions and which questions. Even though the corresponding LF representations for the two questions are not the same, both the LF2 answer to a how many question and the single answer to a which question involve the intersection of all sets mentioned in the story.

### 4.1 Method

**Subjects** Thirty two undergraduate students from the University of Massachusetts at Amherst completed the questionnaire as part of a half hour experiment (six short questionnaires from different studies were to be completed) and received extra course credit for it.

**Materials** There were four versions of each story/how many question pair and two versions of each filler story/which question pair. Example items are provided in Appendix 1.

**Procedure** Participants were presented with a two-page questionnaire. The top of the first page instructed them on the task, telling them that they should read through the stories and answer the questions with the first answer that came to mind. From the answer it was usually possible to infer the interpretation that was associated with the question. They were also asked to mention if there was more than one possible answer.

The questionnaires contained eight stories each followed by an ambiguous how many question and four filler stories followed by an ambiguous which question. Across the four forms of the questionnaire, the questions were counterbalanced with respect to the type of how many constituent (partitive versus non-partitive), and with respect to the order in which the relevant cardinality information appeared in the text (order 1 versus order 2). The filler stories were also counterbalanced with respect to the order of the information. The four questionnaires were randomized once only.

### 4.1.2 Results

Contrary to the predictions of the Minimal Cost Hypothesis, the how many questions were answered more often with an LF1 answer (58.3% of the times) than with an LF2 answer (37.8% of the times). The full data appear in Table 1. There was a significant preference for the LF1 answer over the LF2 answer (i.e. freq(LF1)-freq(LF2) > 0), by subject ($t_{1(31)} = 2.75$, $p < 0.007$) but not by item.
Table 1  English *how many* questions (number of answers)

<table>
<thead>
<tr>
<th>Condition</th>
<th>LF1 answer</th>
<th>LF2 answer</th>
<th>Cumulative answer</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. non-partitive/order1</td>
<td>39</td>
<td>22</td>
<td>3</td>
<td>64</td>
</tr>
<tr>
<td>b. non-partitive/order 2</td>
<td>35</td>
<td>24</td>
<td>4</td>
<td>63</td>
</tr>
<tr>
<td>c. partitive/order1</td>
<td>36</td>
<td>25</td>
<td>2</td>
<td>63</td>
</tr>
<tr>
<td>d. partitive/order 2</td>
<td>38</td>
<td>25</td>
<td>1</td>
<td>64</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>148 (58.3%)</strong></td>
<td><strong>96 (37.8%)</strong></td>
<td><strong>10</strong></td>
<td><strong>254</strong></td>
</tr>
</tbody>
</table>

*a* order 1 = in the story, LF1-information appears before LF2-information.

*b* order 2 = in the story, LF2-information appears before LF1-information.

Table 2  English *which* questions (number of answers)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Single answer</th>
<th>Pair-list answer</th>
<th>Cumulative answer</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>order 1</td>
<td>60</td>
<td>2</td>
<td>2</td>
<td>64</td>
</tr>
<tr>
<td>order 2</td>
<td>61</td>
<td>1</td>
<td>1</td>
<td>63</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>121</strong></td>
<td><strong>3</strong></td>
<td><strong>3</strong></td>
<td><strong>127</strong></td>
</tr>
</tbody>
</table>

(t2(7) = 1.59, p < 0.15). However, seven out of the eight stories had a significant preference for the LF1 answer (sign test p < 0.035).

There were also ten cumulative answers (answers that count the total number of objects in the story) and some second choice answers (as a second choice there were four LF1 answers, two LF2 answers and one cumulative answer). These answers were not included in the statistical analysis. Two answers were invalid since they didn’t correspond to any possible answer and could thus not be interpreted.

As can be seen in Table 1, there was no overall difference between the two differently-ordered versions of each story. Hence, the preference for the LF1 answer was independent from the order in which the information appeared in the text. Furthermore, there was almost no difference between partitive and non-partitive questions.

Concerning the filler items, there was a very strong preference to give a single answer to a *which* question (t1(31) = 20.5, p < 0.0001, t2(3) = 13.8, p < 0.0009). Note that the single answer involves picking out the same set as would a corresponding LF2 answer. Even though subjects dispreferred an LF2 answer to a *how many* question, they did not disprefer the corresponding single answer to a *which* question. Here as well, the order in which the information appeared in the text had no effect. The results are presented in Table 2.
4.1.3 Discussion  The results of this experiment indicate that participants preferred to choose an LF1 answer over an LF2 answer, when asked to respond to an ambiguous how many question. We can conclude that the Minimal Cost Hypothesis does not make the correct predictions. Perceivers seem to prefer to interpret how many questions with an LF1 interpretation. These results thus suggest that the parsing mechanism cannot be guided by the kind of Economy considerations expressed in the Minimal Cost Hypothesis. At least, these considerations are not sufficient to determine which LF is to be associated with a sentence.

What other considerations could play a role in the resolution of quantifier scope ambiguities? At first sight, one might speculate that the LF2 answer demands more cognitive computations than the LF1 answer, since it requires determining the intersection set of all sets mentioned in the story. However, since the which questions showed a very strong preference for a single answer, such a proposal cannot be maintained. The single answer picks out exactly that intersection set. Since it could be claimed that pair-list answers might have been avoided out of ‘laziness’ (participants could be avoiding to write out the rather long answer), I constructed a similar follow up questionnaire in which 16 participants were given stories followed by a question and two possible paraphrases for the answer. They were asked to choose one of the two paraphrases. In this study, there was still a strong preference to respond to a which question with a single answer (50 single answers and only 14 pair-list answers). One can thus reject the hypothesis that the LF2 answers were avoided because computing the intersection set would be too demanding.

Interestingly, none of the scope principles used in the linguistic and psycholinguistic literature predict the preference for an LF1 answer to a how many question. For the case of declarative sentences, Kurtzman & MacDonald (1993) provide experimental evidence that a number of scope principles interact. They argue for a strong influence of two factors on scope preferences in declarative sentences. One of these favours either a linear left-to-right hierarchy or a structural high-to-low ordering (the linear order principle, Bunt 1985; Fodor 1982; and many others, and the c-command principle, Reinhart 1983). The other factor favours an ordering in terms of thematic hierarchy or a preference to give external arguments higher scope (the surface subject principle, ioup 1975, and the thematic hierarchy principle, Grimshaw 1990; Jackendoff 1972). Kurtzman & McDonal argue that in the case of active declarative sentences both factors predict a preference for the leftward quantified phrase to take wide scope, as evidenced by their
experimental results. In the case of passive declarative sentences, the factors conflict since in a passive sentence the subject does not precede the object. Here, no clear preference is predicted, again as evidenced by their results.

If scope resolution in questions is to result from these same scope principles, these are yet to be defined in a way that they can capture all the possible scope construals of a question. In particular, the present scope principles do not capture the fact that wh-elements may contain more than one quantificational element, as is the case for how many N (For which n + n-many N), and that another quantificational element might take scope between these two elements, as in LF1 (order: For which n, every N, n-many N). Putting aside this issue, one could propose that the scope principles ignore the interrogative quantifier For which n and only control the preference of order for every N and n-many N. Interestingly, the two factors presented above conflict in this case: the c-command principle and linear order principle predict a preference for the surface structure order, hence an LF2 representation, while the thematic hierarchy principle and surface subject principle predict a preference for the subject to take scope over the object, hence an LF1 representation. Following Kurtzman & MacDonald’s proposal, in analogy to passive declarative sentences, we then expect no particular preference for one or the other interpretation of a how many question. This is contrary to what has been observed in Experiment I.

Finally, the Principle of Scope Interpretation proposed by Tunstall (1997) does not make any predictions about this case since it says nothing about the cost of different landing sites of quantifiers in the LF representation of a question. The Principle of Scope Interpretation states that the parser chooses to construct the LF that requires the minimum of necessary grammatical operations, or in other words, the LF that only requires grammatically necessary movements. Notice that, contrary to the results presented here, we would expect a preference for the LF2 representation over the LF1 representation. Under this principle one would expect the lowering movement of the quantifier n-many N to be as short as possible: in order to construct an LF representation that is interpretable it is sufficient to lower the quantifier below the question operator and not farther. This results in an LF2 representation. Nevertheless, the results of Experiment I have shown that an LF1 representation is in fact preferred.

Before proposing an analysis for the results of this experiment, I turn to the French data which present additional evidence against the Minimal Cost Hypothesis.
5 FURTHER EVIDENCE FROM FRENCH HOW MANY QUESTIONS

After having discussed ambiguous how many questions in English in the previous section, I will now turn to a kind of how many construction in French that does not display this ambiguity. French makes use of a corresponding split construction, in which only combien (‘how many’) is fronted. The important feature of this construction is that it can only be interpreted with an LF1 interpretation and not with an LF2 interpretation. For convenience, I repeat the scenario from Section 2 in (11). The question in (12) can only receive the ‘six papers’ answer in this scenario.

(11) In the music department, three trumpet students had to pass an exam last week. Every student had to play six pieces. The only requirement they had was that among these there were two that everybody had to play: ‘Round Midnight’ and ‘The days of Wine and Roses’. For the rest, they were free to choose what they preferred.

(12) Combien tous les étudiants ont-ils joué de pieces?
   How many all the students have-they played of pieces?
   How many pieces did every student play?

   a. √ Answer 1: Six pieces.
   b. * Answer 2: Two pieces. 4

The French non-split counterpart, however, behaves like the English construction in that it is ambiguous and permits both answers, as illustrated in (13).

(13) Combien de pieces tous les étudiants ont-ils joués?
   How many of pieces all the students have-they played?
   How many pieces did every student play?

   a. √ Answer 1: Six pieces
   b. √ Answer 2: Two pieces

The underlying LF representations for the French non-split construction are assumed to be identical to the LF representations that have been proposed for English in section 2. For convenience, I repeat these LF representations in (14).

4 The asterisk indicates that this is an impossible answer.
French thus presents the case where two different constructions can be used to express the same interpretation (the LF1 interpretation). While the split construction can only be used to express that particular reading, the non-split construction is ambiguous.

The Minimal Cost Hypothesis presented above predicts the preference for an LF2 representation of a how many question. In the following, I argue that there is independent reason to expect this preference for the non-split question in French.

Because French admits two constructions that differ with respect to their possible scope configurations, a particular version of the Blocking Principle (cf. Aronoff 1976) can be argued to apply. The Blocking Principle states that the existence of a particular word blocks the existence of another word with identical semantics. Di Sciullo & Williams (1987) and Williams (1996, 1997) propose that this principle can also be extended to other levels of the grammar, in particular to syntax. Williams (1997) interprets Aronoff’s (1976) principle in the following way: ‘if two forms exist (in syntax or morphology), they must have different meanings’ (p. 578).

The question whether the effects of the Blocking Principle should extend beyond the area of morphology is still under debate. Note that the French configuration, in fact, disconfirms Williams’ (1997) proposal. That proposal predicts that the non-split construction should not give rise to ambiguity at all. Williams (1997) claims that the Blocking Principle predicts the following two possibilities: ‘either one form is associated with the meaning, in which case the other form cannot have the meaning (blocking); or there is only one form which is ambiguous. If there are two forms, they must have different meanings; if there is only one form, it is permitted to be ambiguous.’ (p. 585).

However, there exists a pragmatic version of the Blocking Principle which follows directly from the Gricean principles (cf. Grice 1975). The principle against obscurity requires that a question be phrased as unambiguously as possible. Consequently, in the case of French, we expect the non-split construction to be used to express an LF2 interpretation. If the questioner wanted to convey the LF1 reading of the question, the principle against obscurity would force her/him to
chose an unambiguous construction instead of the ambiguous non-split construction. The split construction should then be used whenever the LF1 interpretation is the intended interpretation of the question.

Under a Gricean version of the Blocking Principle, we thus expect the non-split construction to be used to express the LF2 interpretation rather than the LF1 interpretation. Consequently, the parser should prefer to assign the LF2 representation to the non-split construction. The hypothesis that makes this prediction is stated in (15).

(15) Blocking Hypothesis

In French, the split question is unambiguously used to express the LF1 interpretation. Assuming a Gricean version of the Blocking Principle, we expect the non-split question to be used in production to express the LF2 interpretation, and correspondingly expect perceivers to favour that reading in comprehension.

Clearly then, for French, the Minimal Cost Hypothesis in (8) and the Blocking Hypothesis in (15) together strongly predict a preference to answer a non-split question with an LF2 answer. Experiment II tests these hypotheses.

5.1 Experiment II

This experiment consisted of a French questionnaire study that tested the predictions of both the Minimal Cost Hypothesis laid out in (8) and the Blocking Hypothesis in (15).

As in the previous study, in Experiment I, participants had to answer how many questions preceded by short stories. The stories were constructed exactly as in Experiment I (see Experiment I for details).

Two factors were manipulated: the order in which the information about the number was presented in the text, and the split/non-split nature of the wh-element. As before, the first factor was introduced to eliminate the possibility that participants would simply answer with the last information from the text. The second factor was introduced to test the Blocking Hypothesis.

5.1.1 Method

Subjects Sixty-four undergraduate students from the Université Paris 8 completed the questionnaire as part of a homework assignment in an introductory course to linguistics.
Table 3 French *how many* questions (number of answers)

<table>
<thead>
<tr>
<th>Condition</th>
<th>LF1 answer</th>
<th>LF2 answer</th>
<th>Cumulative answer</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>split/order 1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>103</td>
<td>8</td>
<td>15</td>
<td>126</td>
</tr>
<tr>
<td>split/order 2&lt;sup&gt;b&lt;/sup&gt;</td>
<td>105</td>
<td>8</td>
<td>15</td>
<td>128</td>
</tr>
<tr>
<td>non-split/order 1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>102</td>
<td>14</td>
<td>11</td>
<td>127</td>
</tr>
<tr>
<td>non-split/order 2&lt;sup&gt;b&lt;/sup&gt;</td>
<td>100</td>
<td>8</td>
<td>19</td>
<td>127</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>410 (80.7 %)</td>
<td>38 (7.4%)</td>
<td>60</td>
<td>508</td>
</tr>
</tbody>
</table>

<sup>a</sup> order 1: In the story, LF1-information appears before LF2-information.
<sup>b</sup> order 2: In the story, LF2-information appears before LF1-information.

Materials There were four versions of each story-question pair and two versions of each filler story-question pair. Example items are provided in Appendix 2.

Procedure Participants were presented with a two-page questionnaire. The top of the first page instructed them on the task, telling them that they should read through the stories and answer the questions with the first answer that came to mind. They were also asked to mention if there was more than one possible answer. From the answer it was usually possible to infer the interpretation that was associated with the question.

The questionnaires contained eight stories, each followed by an ambiguous *how many* question, and four filler stories followed by an ambiguous *which* question. Across the four forms of the questionnaire, the questions were counterbalanced with respect to the type of *how many*-constituent (split versus non-split), and with respect to the order in which the relevant cardinality-information appeared in the text (order 1 versus order 2). The filler stories were also counterbalanced with respect to the order of the information. The four questionnaires were randomized once only.

5.1.2 Results French *how many* questions were answered more often with an LF1 answer (80.7% of the times) than with an LF2 answer (7.4% of the times). The full data appear in Table 3. In the case of a non-split question, contrary to the predictions of the Minimal Cost Hypothesis and the Blocking Hypothesis, there was a highly significant preference for the LF1 answer over the LF2 answer by subject ($t_1(63) = 16.32, p < 0.0001$) and by item ($t_2(7) = 12.71, p < 0.0001$).

There were also sixty cumulative answers (answers that count the
total number of objects) and several second choice answers (as a second choice there were two LF1 answers, nine LF2 answers and eighteen cumulative answers). These answers were not included in the statistical analysis. Four answers were invalid since they didn’t correspond to any possible answer and could thus not be interpreted.

As can be seen in Table 3, there was no overall difference between the two differently ordered versions of each story. The preference for the LF1 answer was independent from the order in which the cardinality information appeared in the text. Furthermore, contrary to the Blocking Hypothesis, there was almost no difference between split and non-split questions. The non-split questions received only a few more LF2 answers than the split questions.

Concerning the filler items, there was a very strong preference to give a single answer to a which-question, instead of a pair-list answer $(t_1(63) = 6.82, p < 0.0001, t_2(3) = 5, p < 0.02)$. Here as well, the order in which the information appeared in the text had no effect. There were only a few more single answers for order 2. The results are presented in Table 4.

5.1.3 Discussion The French results provide additional evidence against the Minimal Cost Hypothesis and the Blocking Hypothesis. Contrary to the prediction of these hypotheses, there is a preference to answer a how many question with an LF1 answer rather than with an LF2 answer. The preference for an LF1 answer is in French even stronger than it is in English.5

The results from the which questions closely parallel the English

5 Dekydtspotter et al. (2001) ran a similar experiment with French native speakers and obtained results that closely parallel these. In their experiment, rather than providing an answer to a question, subjects had to judge whether the presented answer was a correct answer to a question in a particular context. 79.22% LF1 answers to split questions were judged correct, and 68.83% LF1 answers to non-slip questions were judged correct. Only 11.69% LF2 answers to split questions were judged correct, and 29% LF2 answers to non-split questions were judged correct. See Dekydtspotter et al. (2001) for details of their experiment.
results. A single answer is preferred over a pair-list answer. Again, this indicates that the LF2 answer cannot have been avoided because it would require more effort when computing the intersection set.

Contrary to the predictions of the Blocking Hypothesis, there was almost no difference between split and non-split questions in French. This suggests that other factors may have priority and appear to override this principle. We will see later how this result can be integrated into the proposal to be developed below.

To conclude, the results from Experiments I and II provide strong evidence against the Blocking Hypothesis and the Minimal Cost Hypothesis. We can conclude that the process of quantifier scope resolution cannot solely be based on the kind of Economy principles that are encoded in the Minimal Cost Hypothesis. Additional factors must be at play when the parser constructs an interpretation for a syntactic structure.

In the following, I develop a proposal in which the attested preferences are derived under the assumption that context plays a determining role in the process of constructing an LF representation for a how many question.

6 THE PROPOSAL: RESOLUTION OF QUANTIFIER SCOPE AMBIGUITIES DETERMINED BY INTERACTION WITH CONTEXT

In this section, I argue that the resolution of quantifier scope ambiguities requires an incremental context-interactive model. That is, the mechanism that interprets syntactic structures computes the meaning of a sentence piece by piece as the constituents are processed on-line. Furthermore, in this process, the mechanism crucially makes use of information provided by the context. An important feature of the mechanism proposed here is that early consultation of context does not necessarily trigger immediate interpretation of a constituent. I will argue that such interaction can in fact also delay interpretation.

That the semantic context plays a crucial role in the resolution of structural ambiguities is, of course, not a novel proposal. On-line context interactive approaches (Marlson-Wilson & Tyler 1980; Crain & Steedman 1985; Altmann & Steedman 1988; and many others) have argued that the semantic context necessarily affects structural ambiguity resolution. Altmann & Steedman (1988), for example, argue for ‘an architecture in which alternative analyses are initially offered in parallel, and are then discriminated among by immediate appeal to the comprehension process under a selective or ‘weak’ interaction’
In their model, the interpretation of an utterance is computed immediately and in an incremental fashion. Additionally, the processing of an utterance is claimed to always be conducted with immediate reference to the discourse context in which it occurs. Crain & Steedman (1985), for example, argue that ‘the primary responsibility for the resolution of local syntactic ambiguities in natural language rests not with structural mechanisms, but rather with immediate, almost word-by-word interaction with semantics and reference to the context’ (p. 321).

At first sight, the framework adopted in generative grammar seems incompatible with incremental semantic processing. In generative grammar, it is usually assumed that syntax is an autonomous module and that semantic interpretation takes place ‘after’ a syntactic structure has been constructed. This kind of model would seem to preclude incremental interpretation. Nevertheless, it is possible to show that this framework is compatible with a model in which interpretation is calculated incrementally. One of the basic principles of the semantic framework that has influenced most work in recent semantic research in generative grammar, namely Montague grammar,6 states that for each syntactic rule there exists a parallel semantic rule and that semantic composition obeys compositionality. Semantic denotations are limited to functions and arguments that can be combined with a very limited set of rules. The combination of these elements necessarily mirrors the underlying syntactic structure. Crucially, each semantic lexical entry of a functor specifies the arguments that it combines with as well as the outcome of functional application. Hence, in such a model it is not problematic to conceive of a mechanism in which partial semantic representations are assigned to incomplete sentences: at each stage of the semantic derivation we have, in principle, the necessary information about what type of arguments are still needed/expected to construct a complete sentence meaning.7 We can then assume that when the parser constructs an LF representation, the semantic module immediately attempts to combine the semantic denotations of the constituents (as

---

6 See Partee (1996) for an overview of the recent developments in semantics initiated and influenced by Montague’s work.

7 I will not elaborate here on the semantic component that computes the partial semantic representations. See Kempson et al. (2000) for a model of incremental semantic processing. Their parsing model is a goal directed deduction process in which the string is processed left to right. Each word of the input projects data (a declarative unit whose formula part represents the word’s type and whose label may contain syntactical features, control and semantic information) and a set of expectations (the declarative units that are expected to occur in the data generated by preceding and subsequent words). This is equivalent to the mechanism outlined here. See Appendix 4 for the semantic representations that can be associated with the LF representations.
soon as they are processed) to construct a partial interpretation of the sentence.

Note, however, that under the assumption that semantic interpretation is calculated in an incremental and immediate fashion and that semantic composition mirrors the underlying syntactic structure, we expect perceivers to prefer an LF2 representation of a how many question. Following Altmann & Steedman (1988), in an interactive processor, each constituent is interpreted as soon as it is encountered. We thus expect the processing mechanism to attempt to incorporate constituents into the LF representation as soon as they are encountered and to combine the corresponding semantic denotations immediately. This assumption is in line with what is proposed in Crocker (1996), who argues that ‘we don’t simply recover the semantic referents of each individual word, but we also rapidly assign grammatical structure to the input, as words are encountered, so that compositional semantic interpretation can take place immediately and incrementally’ (p. 4). It follows that, under the assumption of immediate interpretation, the LF2 representation should be the representation that is constructed for a how many question. The parser should commit to the LF2 representation (and to its corresponding interpretation) as soon as the how many constituent has been encountered, and the subject quantifier every N should be interpreted in the scope of n-many N as dictated by surface structure. Nevertheless, as we have seen, this does not correspond to the preferred reading of a how many question. For convenience, I repeat again the two LF representations in (16a) and (16b).

(16) a. LF1: [CP How_n [Q [IP every student_i [tn-many pieces_j [VP_t played_tj]]]]]  
   LF1 answer: Six pieces

b. LF2: [CP How_n [Q[IP [tn-many pieces_j [IP every student_i [VP_t played_tj]]]]]]  
   LF2 answer: Two pieces

We can conclude that an incremental mechanism as described above, in which constituents are interpreted as soon as they are encountered and in which semantic composition mirrors the underlying syntactic structure, is incompatible with the results of the experiments presented previously. Experiments I and II showed that the LF1 representation was preferred, not the LF2 representation as would be expected.

In the following, I propose an incremental context-interactive model in which the interpretive mechanism is conceived in a slightly different way.
In this model, the parser can access information from the discourse context when constructing an LF representation and its corresponding interpretation. Such a discourse model should be viewed as a conversational record. Following Stalnaker (1979) and many others, sentences are used in communication to contribute to an already existing conversational record, which contains a set of common background assumptions built up among conversational participants. As proposed in Karttunen (1976) the conversational record can be described as a file that consists of records of all the individuals mentioned in the text, and for each individual of a record that contains its properties. More recently, formal discourse models have developed (cf. for example, Discourse Representation Theory, Kamp & Reyle 1993, and File Change Semantics, Heim 1982) in which the meaning of a sentence is seen as a process of updating the conversational record. Indefinites introduce new discourse referents into the representation, while definites and pronouns refer back to an antecedent introduced earlier into the discourse (and add properties to those).

In my proposal, I do not go into the technical details of an incremental semantic framework. Rather, I focus on the conditions/principles that guide the construction of the LF representations corresponding to the interpretations that are at stake. In line with the theoretical work on discourse models discussed above, I argue that the construction of an LF representation is guided by the context when the LF contains anaphoric constituents. Such constituents require an antecedent in the discourse context in order to receive an interpretation.

A key aspect of my proposal is that interrogative constituents (so called wh-elements), like pronouns, must find an antecedent in the context. As for pronouns, it is well known that they require an antecedent in the context in order to be interpreted. One can assume that processing a pronoun triggers a search for an antecedent in the discourse context. A parallelism between pronouns and wh-elements in their discourse properties can be established easily. In the case of wh-elements, the listener has to determine the corresponding antecedent in the context, in order to respond to the question. I thus propose that, similar to the case of pronouns, once a wh-element is perceived it triggers the search for an antecedent in the discourse context. For example, What N requires the listener to find an antecedent (of type N) in the context, and How many N requires the listener to find an antecedent-set (of elements of type N) in order to determine its cardinality. Hence, wh-elements differ from pronouns only in that the search for an antecedent is not done in order to update the context, but
in order to retrieve information from it. Note that this view of wh-elements is compatible with the linguistic literature on quantifiers. It has widely been argued that quantifiers are anaphoric expressions that need to be interpreted with respect to a contextually determined set (see Kratzer 1978; Westerståhl 1985, and more recently van Deemter 1992; von Fintel 1994; Geurts & van der Sandt 1999; Hendriks & de Hoop 2001; Krahmer & van Deemter 1998; Roberts 1995; Stanley 2002; Stanley & Szabó 2000, among others).

The attested scope preferences for how many questions in experiment I and II can now be derived from these anaphoric discourse properties of wh-elements in a rather straightforward manner. In a system in which meaning is calculated incrementally as described above, we expect the search for an antecedent of an anaphoric element to be done as soon as possible, that is, as soon as the constituent is encountered.8 Interpreting the constituent how many pieces amounts to searching for a set of pieces in the discourse in order to determine its cardinality. Note, however, that the wh-phrase by itself does not necessarily have enough descriptive content to allow the parser to choose an appropriate set from the discourse. If there is more than one set of pieces available in the discourse, the parser cannot choose an antecedent immediately. Only if the context contains a unique antecedent is the search successful and an interpretation can be assigned immediately to the constituent. Thus, the constituent can only be interpreted successfully if appropriate discourse conditions are satisfied.

Given this, it seems reasonable to propose that the failure to find a unique discourse antecedent for the constituent n-many N triggers the parser to delay the incorporation of this element into the LF representation (and into the corresponding semantic representation). The parser does not have enough information in order to determine the appropriate antecedent. Under these circumstances, the constituent n-many N is kept in memory. Since immediate interpretation is not possible at that point, the structural incorporation of the element into the LF representation can only be determined later, when more information is available. This is stated in the Context Dependence Hypothesis in (17).

---

8 Experimental evidence for this claim in the case of pronouns is presented in Garrod (1994), who argues that pronouns can be integrated early in sentence interpretation because they depend heavily on context, and are typically used when the discourse entities they refer to are highly active in working memory. The author further suggests that any delay in the interpretation of a pronoun would require ‘holding an uninterpreted place marker’ in working memory, since pronouns contain little lexical information to otherwise aid in their interpretation. This is associated with an added processing cost.
How many N triggers the search for a set in the discourse. A context that provides more than one possible antecedent for n-many N delays the incorporation of this constituent into the LF representation.

Under this hypothesis, immediate access to context does not commit the parser to immediate incorporation of the constituent into the LF representation. Immediate incorporation is only possible when there is one unique (salient) antecedent in the context. Interestingly, similar proposals have been developed for anaphora resolution in the case of pronouns. In the psycholinguistic literature, it has been claimed that early commitment to anaphora resolution is only possible when a pronoun uniquely identifies an antecedent in the context. In any other case, there is evidence for delayed resolution of the sentence (cf. Garrod & Sanford 1994 and references cited therein). Similarly, in the linguistic literature, it has been shown that the use of an anaphoric element is only acceptable, if it is clear which of the discourse referents in the discourse context it refers to (cf. Heim 1982, Kadmon’s 1987 general uniqueness condition on definite descriptions and pronouns).

Importantly, in the experiments presented above, questions were interpreted in contexts that contained more than one possible antecedent set. In these experiments, the scenarios were constructed in a way that neither the LF1 nor the LF2 information would be more salient. For these experiments, then, the Context Dependence Hypothesis does not predict an immediate commitment to the LF2 interpretation (the LF in which n-many N is integrated and interpreted as soon as it is encountered). Rather, the Context Dependence Hypothesis predicts that the incorporation of the n-many N constituent into the LF representation should be delayed.

How can the attested preference for an LF1 interpretation be captured by the proposed model? We will see that this preference is captured under the assumption that two principles play a role in the processing of a how many question: Immediate Interpretation and the Structure Preservation Principle.

Given the Principle of Immediate Interpretation, the parser attempts to combine the constituents that it encounters to form a partial semantic representation as soon as they are processed.9 Consider the parser at the stage after the decision to not integrate n-many N into the LF representation. Under the pressure toward immediate incremental

---

9 See Appendix 4 for the underlying semantic representations of the incomplete LF structures discussed here.
analysis, the parser attempts to integrate the constituents that follow into the LF representation, even though the *n*-many *N* constituent is kept in memory for the moment. At this stage, the parser can thus construct the following structure.

(18) \([CP \text{ How}_n [Q [IP every student]]]\)

This procedure is in line with Crocker’s (1996) Principle of Incremental Comprehension. Crocker argues that ‘the sentence processor operates in such a way as to maximize comprehension of the sentence at each stage of processing’ (p. 106). He claims that under Immediate Interpretation ‘any structure which can be built, must be’ (p. 107).

After processing the subject quantifier *every N*, the verb is encountered. At this point, a decision has to be taken as to what to do with the constituent *n*-many *N* that has been kept in memory so far. Incorporating the verb into the LF representation before incorporating the *n*-many *N* constituent would lead to an illicit semantic representation. In order for the LF representation to receive a licit interpretation, the quantifiers have to be in a structurally higher position than the Verb Phrase. Consequently, the constituent *n*-many *N* must be incorporated into the LF representation once the verb has been encountered. At that point, the parser thus has to decide about the order of the quantifiers in the LF representation. Crucially, however, the information given by the verb is not sufficient to disambiguate between the LF1 information and the LF2 information, since the question as a whole is ambiguous.

At this stage, the Structure Preservation Principle comes to play a crucial role for the decision whether to construct the LF1 or the LF2 representation. This principle requires that, when building LF representations, the parser should give up as little successfully built and interpreted structure as possible (cf. the Minimal Revision Principle by Frazier 1990 which requires that the parser, when making revisions, maintain as much of the already assigned structure and interpretation as possible). Since, at the point at which the parser has to decide about the order of the quantifiers, the subject quantifier has already been successfully integrated into the LF representation (cf. (18)), *n*-many *N* is integrated into the LF representation below the structure that has already been built, and thus ends up in a structurally lower position than the subject quantifier (i.e. in the scope of the subject quantifier), as in (19).

(19) \([CP \text{ How}_n [Q [IP every student] [n-many pieces]]]\)
The resulting LF representation is as in (20), in which the Verb Phrase has been incorporated.

(20) LF1: \[ CP \text{ How}_n \ [ Q \ [ IP \text{ every student}_i \ [ [n\text{-many pieces}_j \ [ VP_t \text{ played}_t ]_j ]_j ]_j ]_j ]_j ]

Under the Context Dependence Hypothesis in interaction with Immediate Interpretation and Structure Preservation, we can thus derive the preference for an LF1 representation observed in Experiments I and II.

Under what circumstances can we expect a preference for an LF2 interpretation? Contexts that contain a unique antecedent for \( n\text{-many} \ N \), or at least a unique particularly salient antecedent, can lead to the construction of an LF2 representation. In such a context, the parser can immediately assign an interpretation to the constituent and integrate it into the LF representation because its antecedent can unambiguously be identified. Hence, a commitment to the LF2 representation results. For convenience, I repeat again the LF2 representation in (21).

(21) LF2: \[ CP \text{ How}_n \ [ Q \ [ IP \text{ [n\text{-many pieces}_j ]_j } \ [ IP\text{ every student}_i \ [ VP_t \text{ played}_t ]_j ]_j ]_j ]_j ]

In order to test the Context Dependence Hypothesis I conducted a self-paced reading study in English that allowed me to measure reading times for individual regions of a question. The Context Dependence Hypothesis predicts that the type of context should influence the processing of a \textit{how many} question. In particular, in contexts that contain no unique salient antecedent, we can expect a higher processing load, and thus longer reading times, at the point when \( n\text{-many} \ N \) is integrated into the LF representation (once the verb has been encountered). In this experiment, the contexts of the previous experiments were compared to contexts in which the antecedent-set corresponding to the LF2 answer was made particularly salient.

6.1 Experiment III

The Context Dependence Hypothesis states that the processing of a \textit{how many} question is dependent on the type of context that precedes the question. It predicts that contexts that do not provide a unique salient antecedent-set for \( n\text{-many} \ N \) have a higher processing cost once the verb has been encountered, at the point when the \( n\text{-many} \ N \) constituent is integrated into the LF representation. Experiment III was designed to test this hypothesis.
In order to measure this processing difficulty, the self-paced reading method was employed. Participants had to read story/how many question pairs on a computer screen, and were asked to choose one out of two possible answers presented on the screen (LF1 answer and LF2 answer).

The materials were taken from the English questionnaire study described in Experiment I. For each of the stories from Experiment I, which were meant to support both interpretations equally, a minimally different version was constructed. This second version was meant to support more strongly the LF2 answer. The stories were modified in the following way. In order to increase the salience of the set corresponding to the LF2 interpretation, at the end of each story, one or two sentences were added that involved this set in one additional event. Furthermore, the salience of the information corresponding to the LF1 answer was reduced by omitting the actual cardinality of the corresponding set, or replacing it with a vague cardinality like several, different or some. It was expected that this would further increase the salience of the set corresponding to the LF2 answer. This is illustrated in example (23), the modified version of the original story in (22).

(22) Original story (supports both the LF1 and LF2 answer equally)
In December, the chef distributed some of his recipes to his students.
There was one recipe that everybody received:
the ‘Chilled Terrine with Pistachios and Caper Mustard’.
Altogether, each of them received four different recipes.

(23) Modified story (supports the LF2 answer)
In December, the chef distributed some of his recipes to his students.
There was one recipe that everybody received:
the ‘Chilled Terrine with Pistachios and Caper Mustard’.
That was his special recipe.
He wanted to make sure that everybody would be able to try it out.

Both (22) and (23) were followed by the question in (24).

(24) How many recipes did every student receive from the chef in December?

The factor of context was manipulated in order to see whether the context would have an effect on the processing of the question. In
Condition 1, the original stories from Experiment I were used. I will call this condition the Multiple Sets Condition, because in the original stories the wh-element had multiple antecedent-sets. In Condition 2, the modified versions of these stories were used. I will call this condition the Unique Salient Set Condition, because in the modified stories the wh-element had one unique salient antecedent-set. Since it had been established before that differently ordered stories had the same preferences, this factor of order was not examined anymore. Half of the items were chosen with order 1 (LF1 information before LF2 information) and the other half with order 2 (LF2 information before LF1 information).

The self-paced reading method allowed me to measure the reading times associated with the different regions of a how many question. Higher reading times were expected for the regions that follow the verb in the Multiple Sets Condition than in the Unique Salient Set Condition. Only in the Multiple Sets Condition was an extra processing load expected after that region. At that point, the quantifier still has to be integrated into the LF representation. This increase in processing load for the integration of the quantifier is only expected in the Multiple Sets Condition, since here the quantifier has to be processed and integrated into the LF representation in addition to the constituents that are encountered on-line (in the regions that follow the verb). To the contrary, in the Unique Salient Set Condition the quantifier is integrated into the LF representation immediately when encountered and no other constituents are processed at that time.

Additionally, more LF2 answers were expected in the Unique Salient Set Condition, since here the n-many N constituent can be interpreted immediately and we predict that the LF2 representation should be constructed.

6.1.1 Method

Subjects Fourty-four undergraduate students from the University of Massachusetts participated in this self-paced reading study and received extra course credit for it.

Materials There were two versions of each story–how many question pair. See Appendix 3 for example items.

Procedure Participants read twelve stories, each followed by a how many question, on a computer screen (together with fifty four stories from other studies). The self-paced reading method was employed. To
complete the task participants used a response console that had a left and a right trigger. Either of the two triggers could be pulled to make a phrase appear on the screen. Participants were asked to pull a trigger as soon as they had read through the portion of text that had appeared on the screen. The stories were presented cumulatively, that is, each time when the trigger was pulled and new text appeared on the screen, the preceding portion of text did not disappear. The experiment was designed in this way to make sure that participants would read the presented information as a coherent story. They were told that, at the end of the story, they could read through the whole text again, if necessary. At that point, when the trigger was pulled once again, the story disappeared from the screen and the question was presented. The question, on the contrary, was not presented cumulatively, that is, participants could only see one portion of text at a time by pulling a trigger.

The questions were presented in five separate regions: the how many phrase, the subject quantifier every N, the verb (optionally with a particle) and two modifiers (Prepositional Phrase or Noun Phrase). After all five regions had been presented, the question disappeared and two possible answers appeared to the right and the left of the screen. An answer could be chosen by pulling the corresponding right or left trigger. Participants had been asked to choose the answer that first came to mind. Half of the answers presented the LF1 answer to the left, the other half presented the LF1 answer to the right. For the Unique Salient Set contexts in which the LF1 information was not available (because a vague adjective had been used in the text), the proposed LF1 answer was Don’t know.

The twelve items were counterbalanced for the type of context, constructed as described above (the Multiple Sets Condition versus the Unique Salient Set Condition). The items were randomized separately

---

### Table 5 Reading times in ms for how many questions

<table>
<thead>
<tr>
<th>Condition</th>
<th>Region 1 How many N</th>
<th>Region 2 Quantifier</th>
<th>Region 3 Verb(+Part.)</th>
<th>Region 4 Modifier</th>
<th>Region 5 Modifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple Sets	extsuperscript{a}</td>
<td>539 ms</td>
<td>636 ms</td>
<td>572 ms</td>
<td>713 ms</td>
<td>912 ms</td>
</tr>
<tr>
<td>Unique Salient Set	extsuperscript{b}</td>
<td>541 ms</td>
<td>590 ms</td>
<td>586 ms</td>
<td>640 ms</td>
<td>823 ms</td>
</tr>
</tbody>
</table>

	extsuperscript{a} Multiple Sets Condition = context that supports both interpretations (with no salient antecedent for n-many N).

	extsuperscript{b} Unique Salient Set Condition = context that support the LF2 interpretation (with a unique salient antecedent for n-many N).
for each participant. Because the same experiment included a study in which stories in the future tense were followed by a continuation sentence (instead of a question), three filler stories in the future followed by how many questions were inserted. This was done in order to avoid the possibility that participants could predict from the tense of the story whether they would receive a question or a continuation sentence. These future-stories were not included in the statistical analysis.

6.1.2 Results As predicted by the Context Dependence Hypothesis, in the last two regions (following the verb), participants were significantly slower in the Multiple Sets Condition than in the Unique Salient Set Condition. That is, they were slower when the question was presented in a context that supported both interpretations equally. The two regions following the verb, region 4 and 5, taken together had significantly longer reading times in the Multiple Sets Condition than in the Unique Salient Set Condition (significant by subject, $F_1(1,43) = 5.01$, $p < 0.03$, and nearly significant by item, $F_2(1,11) = 4.59$, $p < 0.06$). The difference in reading times was only significant for the last two regions if taken together, not for each region individually (region 4: $F_1(1,43) = 4.25$, $p < 0.05$, $F_2(1,11) = 3.12$, $p < 0.2$; region 5: $F_1(1,43) = 2.53$, $p < 0.2$, $F_2(1,11) = 3.99$, $p < 0.08$). None of the other regions showed a significant difference. Participants seemed to be slower in the second region, but not significantly so ($F_1(1,43) = 3.06$, $p < 0.09$, $F_2(1,11) = 1.42$, $p < 0.3$). The reading times are given in Table 5 and are represented with a graph in Figure 1.

Furthermore, as expected, the Multiple Sets Condition received significantly more LF1 answers (66% LF1 answers) than the Unique Salient Set Condition (47% LF1 answers), ($F_1(1,43) = 10.13$, $p < 0.003$, $F_2(1,11) = 7.73$, $p < 0.02$).

6.1.3 Discussion The results confirm the Context Dependence Hypothesis. In the last two regions of the question, reading times were higher in the Multiple Sets Condition than in the Unique Salient Set Condition, as reported in Table 5 and Figure 1. This is because in the Multiple Sets Condition, once the verb has been encountered, the quantifier (that is kept in memory) still has to be incorporated into the LF representation. No such operation is necessary in the Unique Salient Set Condition. One can thus explain the higher processing load in the Multiple Sets Condition in the regions following the verb.

As mentioned above, the results were significant only if the last two regions were taken together, not for each region individually.
This is not an undesirable result, since the hypothesis does not make any claims about the exact region in which the decision is taken to integrate the quantifier into the LF representation. In some cases the parser might have waited for further information after the verb in order to disambiguate the two LF representations before integrating the \textit{\textit{n}-many N} constituent into the representation.

The fact that participants were also somewhat slower in the second region in the \textit{Multiple Sets} Condition could be argued to support an alternative hypothesis, namely, that the contexts in the \textit{Unique Salient Set} Condition induced a general lower processing load for the whole question. Since these contexts only supported the LF2 interpretation, and hence were less ambiguous, one could expect the question in the given context to have a general lower processing load. However, if some general difficulty of the context in the \textit{Multiple Sets} Condition slowed reading in both region 2 and region 4 for certain items, then one would expect reading time differences between the \textit{Multiple Sets} Condition and the \textit{Unique Salient Set} Condition in region 2 to correlate across items with the corresponding differences in region 4. No correlation was found (Pearson $r = 0.079$). It can thus be concluded that contexts with a unique salient antecedent did not induce a general lower processing load. Only the last two regions of the question (following the verb) were affected in a significant way.

Finally, the Context Dependence Hypothesis predicts that, in the \textit{Unique Salient Set} Condition, the context ‘guides’ the parser to first attempt to construct the LF2 representation. As a consequence, more LF2 answers were expected in this condition. The results confirm this prediction. The frequency of LF2 answers was significantly higher in the \textit{Unique Salient Set} Condition than in the \textit{Multiple Sets} Condition (and also higher than what was found in Experiment I). These results should be interpreted with caution for the following reason. Since, in the \textit{Unique Salient Set} Condition, the LF1 answer was \textit{Don’t know} (because the cardinality was not available in the text), an alternative explanation of the results is that participants were simply avoiding a ‘Don’t know’ answer, and hence there is an increase in LF2 responses. But there is no support for such an explanation. In fact, it is unlikely that, in the given experimental situation, participants would resort to an avoidance strategy. At the moment when the participant has to answer the question, the text has disappeared from the screen. Participants tend to overlook information on numbers in a self-paced reading task and typically forget the exact numbers. There is evidence that participants do not shy away from responding with a ‘Don’t know’ answer. In the questionnaire studies, participants typically gave a ‘Don’t know’ answer.
when they wanted to answer with a cumulative answer and didn’t have enough information to compute it. Participants would explicitly say that they didn’t know because they didn’t have enough information. This suggests that participants do not necessarily prefer to choose an unintended answer in order to avoid an answer of the kind *Don’t know*. In that sense, the results do confirm the Context Dependence Hypothesis.

To summarize, the results reported in this experiment provide on-line evidence for the Context Dependence Hypothesis. Most importantly, the results indicate that the context that precedes a question plays a crucial role for the resolution of quantifier scope ambiguities. In particular, a context that does not contain a unique salient antecedent set for the wh-phrase delays the resolution of scope ambiguity.

Can the parallel constraint-satisfaction approach advocated in Kurtzman & MacDonald (1993) provide an alternative explanation for these results? As discussed before, in section 4, their principles conflict when applied to questions. In their framework, no preference for one or the other interpretation of the question is expected. Even if one assumed that their principles favouring the LF1 interpretation (the *linear order principle* and the *c-command principle*) were stronger and thus would win over those that favour the LF2 interpretation (the *surface subject principle* and the *thematic hierarchy principle*), it is not clear
whether this model correctly predicts the outcome of Experiment III. In a constraint-satisfaction approach, the difference between the two conditions translates into a difference in competition between alternative interpretations. The Unique Salient Set Condition provides a single clear antecedent for the quantifier, whereas the Multiple Sets Condition does not. This results in more competition between alternative interpretations in the Multiple Sets Condition than in the Unique Salient Set Condition. The constraint-satisfaction approach thus also predicts a delay in interpretation of the question for the Multiple Sets Condition, since more competition is involved. If the principles favouring LF1 are stronger than the principles favouring LF2, we expect the LF1 representation to win over the LF2 representation unless context favours the second. Crucially, the moment at which the parser determines that the context (dis)favours an LF2 representation should be quite early in the sentence, namely when the wh-constituent is encountered. The wh-constituent triggers a search for an antecedent set in the context. In the Multiple Sets Condition, at that point, the decision should be taken that the LF1 is favoured. Hence, in a model in which alternatives compete, one expects the delay caused by the competition to appear immediately after the wh-element has been encountered. In such a model, there seems to be no principled reason to expect the delay much later in the sentence, as it was observed in Experiment III.

The results of this experiment also make a contribution to the debate about what it means for a discourse entity to be ‘salient’. As is well known, diverse factors contribute in making a discourse entity ‘salient’ and it is a difficult task to develop a precise measure of ‘salience’. In this experiment, I identified one factor that provides a measure for the ‘salience’ of a discourse entity: the number of events in which the discourse entity is involved. In the stories, the ‘salient’ set was involved in just one additional event.

Two important questions still need to be addressed. In the questionnaire studies, the results for the how many questions differed strongly from the which questions. Can the proposed model also account for scope ambiguity resolution in which questions? One possibility would be to assume that which and how many do not have the same anaphoric properties. Hence, the Context Dependence Hypothesis would not make any predictions for which phrases. Kluender (1998) indeed proposes that there are important differences between the two types of phrases. He proposes that which phrases are, like definite descriptions, low accessibility markers (following Ariel 1990). They refer to mental referents that are present but not currently active in the discourse model. Under this view, which phrases do not rely on
the context immediately preceding the question. The preference for a single answer can be understood if we assume that a which phrase does not require a unique salient antecedent in the context in order to be interpretable. Instead, a which phrase triggers the postulation of a discourse entity without need for evidence in the context.\(^\text{10}\) Contrary to a how many phrase, a which phrase can thus be integrated immediately into the LF representation even without an appropriate antecedent in the context.

Finally, the questionnaire studies showed that, in French, split and non-split questions did not differ at all in their preference, contrary to the predictions of the Blocking Hypothesis. Can a context interactive model account for this fact? The Context Dependence Hypothesis predicts that a how many phrase will not be incorporated into the LF representation if there is not enough information to determine its antecedent set. Under the Context Dependence Hypothesis, split and non-split questions are, in fact, not expected to have different preferences. In the particular conditions of Experiment II, where contexts supported both interpretations alike, for split as well as for non-split questions, it was impossible to interpret the how many phrase immediately. In the case of split-questions, the full wh-phrase can only be interpreted when all constituents are encountered on surface. In the non-split case, the context delays interpretation of the n-many N phrase until after the verb. A self-paced reading study in French could reveal more about the processing of split and non-split questions. In particular, it could determine whether the moment in which the constituent is integrated into the LF representation differs for split and non-split questions. This will be left for future research.

7 CONCLUSION

The experimental results presented in this paper confirm that there is a preference to interpret a how many question with an LF1 representation (in contexts that do not provide a unique salient antecedent set).

I have argued that this result is unexpected in a system in which the parser obeys economy by first attempting to construct the LF

\(^{10}\) In fact, such a proposal has been made in the literature before. Frazier et al. (1996) propose that which phrases trigger the parser to postulate a discourse entity. If later in the sentence it becomes clear that the which-phrase requires a bound variable interpretation, the parser has to retract the initial assumption and slower reading times result (as in Which rumor about herself did the newspaper claim every actress made up?). These results can be interpreted in the following way: which phrases do not need to access context in order to be interpreted, since they trigger the parser to postulate a discourse entity (even in the absence of evidence for it).
representation that minimally differs from the surface representation of a sentence. Furthermore, I have argued that this result is unexpected under current assumptions about incremental processing of meaning.

The existence of a preference for the LF1 interpretation is even more unexpected in the case of French. A pragmatic version of the Blocking principle leads to the expectation that the split and non-split constructions should give rise to different preferences. This is not what we have observed in Experiment II. Both constructions were preferably interpreted with an LF1 interpretation.

I have proposed a model in which the preference for the LF1 interpretation of a how many question derives from the interaction with context and the assumption that two important principles are at play: Immediate Interpretation and Structure Preservation. On the basis of the anaphoric properties of wh-phrases, I have argued that the interaction with context can be a reason to delay ambiguity resolution and integration of a constituent into the LF representation. Crucially, then, immediate access to context does not necessarily commit the parser to the LF2 interpretation of a how many question. Only contexts that provide a unique salient antecedent trigger immediate interpretation. Thus, in the proposed model, immediate access to context does not necessarily trigger immediate interpretation of a constituent. Under certain conditions, access to context can also delay the interpretation of a constituent (i.e. immediate interpretation is only possible under the appropriate discourse conditions). In this respect, my model crucially differs from previous on-line context interactive models. I have presented results from a self-paced reading study in English that support this model.

Finally, in this paper, I have identified a measure for the ‘salience’ of discourse antecedents. It is well known that discourse ‘salience’ plays an important role in processing. However, it is difficult to develop a precise measure for this notion. The results of experiment III have allowed me to identify one factor that provides a measure for the ‘salience’ of a discourse entity: the number of events in which an entity is involved determine its salience in the discourse. We have thus made some first steps towards a better understanding of the notion of ‘salience’ in semantic processing.

8 ACKNOWLEDGMENTS

I am indebted to Lyn Frazier for invaluable advice throughout every stage of this project. I am grateful to Chuck Clifton for many discussions and helpful comments, and for providing me with the opportunity to
run the experiments in his lab. Furthermore, I would like to thank Gilles Boyé for help with the French questionnaire study and Peggy Speas, Susan Tunstall and Frank Wijnen for useful comments and suggestions. The experimental results of this paper have been presented at three occasions, the 29th North East Linguistics Society Meeting at the University of Delaware in October 1998, the Twelfth Annual CUNY Conference on Human Sentence Processing at the CUNY Graduate School and University Center in March 1999, and in a talk at the Utrecht Institute of Linguistics OTS in November 2000. The paper has greatly benefitted from the comments of the audiences and an anonymous reviewer of the journal. This work was partly supported by NIH grant HD18708 to the University of Massachusetts, Amherst.

ELISABETH VILLALTA
Utrecht Institute of Linguistics—OTS
Trans 10
3512 JK Utrecht
The Netherlands
e-mail: elisabeth.villalta@let.uu.nl

APPENDIX 1: MATERIALS OF THE ENGLISH QUESTIONNAIRE STUDY (EXAMPLE ITEMS)

a. How many question-story pairs

1. Order 1: A group of German tourists spent two days in Manhattan. They were offered different tours which each included a visit to seven sights. All tours included the visit of three sights, the Statue of Liberty, the Empire State Building and Central Park, but otherwise they were very different.
   Order 2: A group of German tourists spent two days in Manhattan. They were offered different tours which all included the visit of three sights, the Statue of Liberty, the Empire State Building and Central Park. For the rest, they were very different. Each tour included a total of seven sights.
   How many (of the) sights in Manhattan did every tour include?

2. Order 1: Three friends went to the last Film Festival in Montreal. Altogether, each of them saw ten movies. When comparing what they had seen at the end, they realized that there were four movies that they all had seen.
   Order 2: Three friends went to the last Film Festival in Montreal. At the end, when comparing what they had seen, they realized that there were
four movies that they all had seen. Altogether, each of them saw ten movies.

How many (of the) movies did everybody see?

3. Order 1: At the conference, Ann, Sarah and Caroline attended ten talks each. They were not interested in the same topics so they did not usually overlap in their choice of talks. There were only three talks that they attended together.

Order 2: At the conference, Ann, Sarah and Caroline attended different talks. They were not interested in the same topics so they did not usually overlap in their choice of talks. There were only three talks that they attended together. During the whole conference, each of them attended ten talks.

How many (of the) talks did everybody attend?

4. Order 1: A group of tourists arrived at the wine fair. Each of them tasted six different wines. The two wines that everybody tasted were the ‘Nouveau Beaujolais’ and the ‘Bordeaux’, because they were praised for being the best wines of the season.

Order 2: A group of tourists arrived at the wine fair. The two wines that everybody tasted were the ‘Nouveau Beaujolais’ and the ‘Bordeaux’, because they were praised for being the best wines of the season. Each of them got to taste six different wines.

How many (of the) wines did every tourist taste?

b. Which question-story pairs

1. Order 1: Recently, Lesley, Gina and Mark realized that they had been on a Europe tour during the same summer. Gina went to London, while Lesley and Mark visited Barcelona. They had all been in the two cities, Paris and Rome, and almost at the same time.

Order 2: Recently, Lesley, Gina and Mark realized that they had been on a Europe tour during the same summer. They had all been in the two cities, Paris and Rome. Gina also went to London, while Lesley and Mark visited Barcelona.

Which cities did everybody visit?

2. Order 1: In the waiting room, Julie, Pat and Liz were reading some of the magazines that were displayed on the table. Julie read PC-complete, Pat the NewYorker and Liz read Emma. All of them read the Newsweek as well.

Order 2: In the waiting room, Julie, Pat and Liz were reading some of the magazines that were displayed on the table. They all read the Newsweek. Julie also read PC-complete, Pat the NewYorker and Liz read Emma.

Which magazines did everybody read?
APPENDIX 2: MATERIALS OF THE FRENCH QUESTIONNAIRE STUDY (EXAMPLE ITEMS)

a. How many question-story pairs

1. Order 1: La semaine dernière, la classe de Matthias devait aller au Futuroscope. Ils ont passé toute la journée sur le site et, en tout, ils ont pu voir six cinémas. Comme il y avait plusieurs groupes, ils ont vu des choses différentes mais ils ont tous pu voir les trois mêmes cinémas: le dynamique, le tapis magique et celui à 3 dimensions.

Order 2: La semaine dernière, la classe de Matthias est allée au Futuroscope. Il y avait plusieurs groupes qui ont vu des choses différentes. Mais ils ont tous pu voir les trois mêmes cinémas : le dynamique, le tapis magique et celui à 3 dimensions. Comme ils ont passé toute la journée sur le site, en tout, ils ont pu voir six cinémas chacun.

Combien de cinémas tous les enfants ont-ils pu voir?
Combien tous les enfants ont-ils pu voir de cinémas?

2. Order 1: Rémi, Elisabeth et Pascale prennent des leçons pour passer le code de la route. Hier, ils ont passé des tests pour simuler l’examen. Ils ont tous fait sept erreurs. Le moniteur leur a expliqué à chacun leurs fautes personnelles. En comparant leurs résultats, ils ont remarqué qu’ils ont tous fait quatre erreurs identiques. C’était sans doute la faute du moniteur...


Combien d’erreurs tous les participants ont-ils fait?
Combien tous les participants ont-ils fait d’erreurs?


Order 2: Un groupe d’amis est parti faire un tour de dégustation à la foire aux vins de Bordeaux. Parmi tous les vins, il y en avait trois que tout le monde avait pu goûter : deux Haut-Médoc et un Graves. Chacun avait pu goûter six vins différents avant d’être ‘indisposé’.

Combien de vins tous les amis ont-ils pu goûter?
Combien tous les amis ont-ils pu goûter de vins?

4. Order 1: Louise, Clara et Aurélie ont chacune une collection de disques assez importante. En tout, elles possèdent une centaine de disques chacune. Elles ont des goûts très différents et cependant il y a une dizaine de disques qu’elles possèdent en commun, les disques de Elvis Presley.
Order 2: Louise, Clara et Aurélie ont chacune une collection de disques assez importante. Elles ont des goûts très différents et cependant il y a une dizaine de disques qu’elles possèdent en commun, les disques de Elvis Presley. En tout, elles possèdent une centaine de disques chacune.

Combien de disques toutes les filles possèdent-elles?
Combien toutes les filles possèdent-elles de disques?

b. Which question–story pairs

1. Order 1: Cette année, au cinéma de plein air, Marie, Fabienne et Sylvain ont vu différents films. Marie a pu voir ‘Le hussard sur le toit’ et ‘La cité de la joie’, Fabienne a vu ‘Les visiteurs’ et ‘La cage aux folles’ et Sylvain a été voir ‘Subway’ et ‘Nikita’. Et, finalement, il y a deux films qu’ils on tous pu voir ensemble: ‘Valmont’ et ‘La vérité si je mens!’

Order 2: Cette année, au cinéma de plein air, Marie, Fabienne et Sylvain ont vu différents films. Il y a deux films qu’ils on tous vu: ‘Valmont’ et ‘La vérité si je mens!’ De plus, Marie a vu ‘Le hussard sur le toit’ et ‘La cité de la joie’, Fabienne a pu voir ‘Les visiteurs’ et ‘La cage aux folles’ et Sylvain a été voir ‘Subway’ et ‘Nikita’.

Quels films tous les amis ont-ils vu?


Order 2: Le chef cuisinier a donné plusieurs recettes à ses trois apprentis. Tout le monde a reçu sa recette de lapin aux pruneaux et celle des éclairs au cresson. De plus, Jean a reçu celle du paté de légumes, Bernard celle des pavés royaux aux raisins secs, et Christine a reçu la recette des courgettes à la grecque.

Quelles recettes tous les apprentis ont-ils reçu ?

APPENDIX 3: MATERIALS OF THE ENGLISH SELF-PACED READING STUDY (EXAMPLE ITEMS)

1. MS Cond.: A group of German tourists spent two days in Manhattan. They were offered different tours which all included a visit to three sights:

   the Statue of Liberty, the Empire State Building and Central Park.

   For the rest, they were very different.

   Each tour included a total of seven sights.

   How many sights did every tour include for the two days in Manhattan?

   Three sights

Seven sights
A group of German tourists spent two days in Manhattan. They were offered different tours which all included a visit to three sights: the Statue of Liberty, the Empire State Building and Central Park. For the rest, the tours were very different. But that didn’t matter so much. Those were the main sights that they wanted to see anyway.

How many sights did every tour include for the two days in Manhattan?

Three sights

Don’t know

At the conference last week, Ann, Sarah and Caroline attended different talks. They were not interested in the same topics. So they did not usually overlap in their choice of talks. There were only three talks that they attended together. During the whole conference, each of them attended ten talks.

How many talks did everybody attend at the conference last week?

Three talks

Ten talks

At the conference last week, Ann, Sarah and Caroline attended different talks. They were not interested in the same topics. So they did not usually overlap in their choice of talks. There were only three talks that they attended together. Those were the talks that they had been asked to report on later. They decided that they would do that together.

How many talks did everybody attend at the conference last week?

Three talks

Don’t know

Last week, a group of my students went to the Metropolitan Museum. Everybody had to visit five galleries and report on them. In particular, everybody had to visit two galleries: The Egyptian Gallery and the Greek Gallery. For the rest, they were free to choose what they wanted to see.

How many galleries did every student have to visit at the Met last week?

Five galleries

Two galleries

The students had to report on the galleries they visited. There were two galleries that everybody had to visit: The Egyptian Gallery and the Greek Gallery. For the rest, they were free to choose what they wanted to visit. These two were also the ones that they had enjoyed most.
How many galleries did every student have to visit at the Met last week?  
Don’t know  
Two galleries

4. MS Cond:  
Mary, Lynn and Cynthia are looking for a new apartment.  
Each of them only really liked three of the apartments they saw.  
Among the apartments they visited, there is one apartment that they all liked.

How many apartments did everybody like among the ones they visited?  
Three apartments  
One apartment

USS Cond:  
Mary, Lynn and Cynthia are looking for a new apartment.  
Among the apartments they liked, there is a particular one they all liked.  
So they will have to fight over it.

How many apartments did everybody like among the ones they visited?  
Don’t know  
One apartment

APPENDIX 4: THE SEMANTIC REPRESENTATIONS CORRESPONDING TO THE LF REPRESENTATIONS

I provide here the semantic representations corresponding to the discussed LF representations, as repeated here.

(1) a. LF1: \[ \lambda p \langle s, t \rangle \exists n \ [ p = \lambda w'. \forall x \text{student}(x)(w') \land \exists Y \text{pieces}(Y)(w') \land \text{card}(Y)(w') = n \Rightarrow \text{play}(Y)(x)(w') ] \]

b. LF 2 reading:
\[ \lambda p \langle s, t \rangle \exists n [ p = \lambda w'. \exists Y \text{pieces}(Y)(w') \land \text{card}(Y)(w') = n \land \forall x \text{student}(x)(w') \Rightarrow \text{play}(Y)(x)(w') ] \]

Adopting a Hamblin/Karttunen semantics (in which a question denotes the set of all possible answers to the question) the semantic representations of LF1 and LF2 are as follows:

(2) a. LF1 reading:
\[ \lambda p \langle s, t \rangle \exists n \ [ p = \lambda w'. \forall x \text{student}(x)(w') \exists Y \text{pieces}(Y)(w') \land \text{card}(Y)(w') = n \Rightarrow \text{play}(Y)(x)(w') ] \]

b. LF 2 reading:
\[ \lambda p \langle s, t \rangle \exists n [ p = \lambda w'. \exists Y \text{pieces}(Y)(w') \land \text{card}(Y)(w') = n \land \forall x \text{student}(x)(w') \Rightarrow \text{play}(Y)(x)(w') ] \]

Following Cresti (1995) and others, the semantic entry of how encodes the quantifier over numbers, and n-many encodes the quantifier over sets. The question operator Q in C turns a proposition into a set of propositions. For simplification, I will ignore the world variable here, and treat propositions p as being of type (t) rather than type (s, t), and thus use an extensional version of a Karttunen-style semantics. The lexical entries for how, n-many and Q are given below.

(3) \[ [[\text{How}]] = \lambda R_{(s, t, t)} \lambda p \exists n \ [ R(n)(p) ] \]
(4) \[[n\text{-}many]\] = \(\lambda Q \langle e, t \rangle \lambda P \langle e, t \rangle \exists Y [P(Y) \& \text{card}(Y)=n \& Q(Y)]\)

(5) \[[Q]\] = \(\lambda q \lambda p [p=q]\)

The bottom-up composition for LF1 is then as follows:

(6) \[[VP]\] = \[[V']\] (y)(x) = played(y)(x)

(7) \[[IP_2]\] = \[[DP_2]\] (\(\lambda y \text{read}(y)(x)\))

(8) \[[n\text{-}many]\] = \(\lambda Q \langle e, t \rangle \lambda P \langle e, t \rangle \exists Y [P(Y) \& \text{card}(Y)=n \& Q(Y)]\)

(9) \[[pieces]\] = \(\lambda y \text{pieces}(y)\)

(10) \[[IP_3]\] = \(\exists Y \text{pieces}(Y) \& \text{card}(Y)=n \& \text{played}(Y)(x)\)

(11) \[[every]\] = \(\lambda Q \langle e, t \rangle \lambda P \langle e, t \rangle [\forall x Q(x) \rightarrow P(x)]\)

(12) \[[CP]\] = \(\lambda p \exists n [p=\forall x \text{student}(x) \rightarrow \exists Y \text{pieces}(Y) \& \text{card}(Y)=n \& \text{played}(Y)(x)]\)

(13) \[[student]\] = \(\lambda y_2 \text{student}(y_2)\)

(14) \[[IP_1]\] = \(\exists Y \text{pieces}(Y) \& \text{card}(Y)=n \& \text{played}(Y)(x)\)

(15) \[[IP_1]\] = \(\forall x \text{student}(x) \rightarrow \exists Y \text{pieces}(Y) \& \text{card}(Y)=n \& \text{played}(Y)(x)\)

(16) \[[Q]\] = \(\lambda q \lambda p [p=\forall x \text{student}(x) \rightarrow \exists Y \text{pieces}(Y) \& \text{card}(Y)=n \& \text{played}(Y)(x)]\)

(17) \[[CP]\] = \(\lambda p \exists n [p=\forall x \text{student}(x) \rightarrow \exists Y \text{pieces}(Y) \& \text{card}(Y)=n \& \text{played}(Y)(x)]\)

I now turn to the semantic representations that can be associated with incomplete syntactic structures. I show that, at each step of the derivation, partial semantic representations can be computed since, for each function, the information on the ‘expected’ argument is available. In order to compute a meaning, such arguments can be postulated.

We begin with the syntactic representation in (21).

(21) \([CP \text{ How}_n [Q]\)

Combining the two semantic entries of \textit{how} and \(Q\) leads to the following partial semantic representation. These two elements combined form a function that requires an argument \(P\) of type \(t\). By postulating this ‘expected’ argument, we have a partial semantic representation of the sentence.

(22) \[[\text{How}][\lambda n.[[[Q]](P)] = \lambda p \exists n [p=R(n)(P)]\]

Next, the parser attempts to integrate the \textit{n}-\textit{many} \(N\) constituent, as in (23).

(23) \([CP \text{ How}_n [Q]_{IP} [n\text{-}many \text{pieces}]\)

This incomplete syntactic representation receives the partial semantic representation as in (24). These elements combined form a function that requires an argument \(R\) of type \(\langle \langle e, t \rangle, t \rangle\). By postulating this argument we have the following semantic representation.

(24) \[[\text{How}][\lambda n.[[[Q]](n\text{-}many)][[[\text{pieces}]](R)] = \lambda p \exists n [p=\exists Y \text{pieces}(Y) \& \text{card}(Y)=n \& R(Y)]\]
This partial semantic representation expresses that the question is about a number n such that there is a set of pieces Y of cardinality n (and this set has an unknown property R). Integrating \( n \)-many \( N \) is only felicitous if the context provides a unique salient antecedent-set. Otherwise, \( n \)-many \( N \) is stored in memory. In this case, continuing to integrate the incoming elements into the syntactic representation leads to the representation in (25).

\[
(25) \quad [CP \ How_n [QIp every \ student,]
\]

This representation receives the partial semantic representation in (26).

\[
(26) \quad ([How][\lambda n.[Q]][[[every]][[[student]]]](R) = \lambda p \exists n [p = [\forall x \ student(x) \rightarrow R(x)]
\]

Notice that in this semantic representation we have vacuous quantification over the variable n. If no other argument is provided, the semantic representation will be illicit. The parser thus expects an argument that will provide the n-argument. Next, the \( n \)-many \( N \) constituent is integrated into the LF representation as in (27).

\[
(27) \quad [CP \ How_n [QIp every \ student, [[tn-many \ pieces]]]
\]

The corresponding partial semantic representation is given in (28).

\[
(28) \quad ([How][\lambda n.[Q]][[[every]][[[student]]]][[[n-many]][[[pieces]]]](R) = \lambda p \exists n [p = [\forall x \ student(x) \rightarrow \exists Y \ pieces(Y) \& card(Y) = n \& R(Y)]
\]

After incorporation of the verb, the final semantic representation of the LF1 reading is the following:

\[
(29) \quad ([How][\lambda n.[Q]][[[every]][[[student]]]][[[n-many]][[[pieces]]]][[[played]]](R) = \lambda p \exists n [p = [\forall x \ student(x) \rightarrow \exists Y \ pieces(Y) \& card(Y) = n \& played(Y)(x)]
\]

REFERENCES


Crain, S. & Steedman, M. (1985) ‘On not being led up the garden path: the use of context by the psychological syntax processor’. In D. Dowty, L. Karttunen, & A. Zwicky (eds), Natural Language Parsing. Cambridge University Press.
Cambridge, UK, 320–358.


