

Seriality in LF Parsing

Anastasia Marie Conroy and Jeffrey Lidz

University of Maryland, College Park

Address:

Department of Linguistics

1401 Marie Mount Hall

University of Maryland

College Park, MD 20742

staceyc@umd.edu

phone: 301-405-7002

fax: 301-405-7104

Abstract

This paper examines the nature of LF parsing with respect to the number of alternative semantic representations considered simultaneously. Much research on phrase structure parsing suggests that multiple phrase structure representations are considered in parallel as the sentence unfolds, with one ultimately chosen as the correct parse. This paper presents a novel experimental procedure, the Incremental Verification Task (IVT), a technique designed to examine the processing of global ambiguity. We find evidence that although the parser can entertain multiple syntactic alternatives simultaneously, LF parsing is fundamentally different. We suggest that the parser processes LF alternatives in serial, first adhering to the representation that most closely matches the surface structure, and revising only if necessary.

keywords: Parsing, Scope Ambiguity

1 Introduction

A key question in the domain of sentence processing is how the parser handles multiple potential analyses of an input string. Since Bever (1970) and Kimball (1973), much attention has been paid to ambiguities that derive from alternative constituent structures (Crain and Steedman, 1985; Clifton and Ferreira, 1989; Frazier and Fodor, 1978; MacDonald, Pearlmutter, and Seidenberg, 1994; Sturt, Pickering, and Crocker, 2000). Multiple models have emerged suggesting that the parser actively pursues multiple representations simultaneously when confronted with potential ambiguity in the phrase structure composition.

Understanding a sentence involves the identification of several kinds of representations. The parser must assign a phonological structure, segment that structure into words, arrange the words into phrases, and assign an interpretation to the words, phrases and entire sentence. A critical question in this domain concerns how phrases are built, with ambiguous sentences often used as a probe into the phrase-structure building component of understanding. From this focus on phrase structure has emerged a consensus (though by no means unanimous) that the parser, when faced with potential ambiguity, actively pursues multiple phrase structure analyses in parallel. It is important to remember, however, that the construction of phrase structure representations is not equivalent to the construction of meaning. A phrase structure must be interpreted. In cases of attachment ambiguities, each possible attachment leads to a

single interpretation and so it is easy to conflate phrase structure building with interpretation building. But these processes are necessarily distinct. The independence of meaning and phrase structure can be seen by examining sentences that have only a single phrase structure analysis but still have two interpretations, such as (1)

- (1) One girl took a bite out of every pie

For clarity, let us refer to the construction of multiple parses for ambiguous strings 'phrase structure parsing'. Let us refer to the construction of multiple interpretations for ambiguous phrase structures (i.e., single phrase structures with multiple interpretations) 'L(ogical) F(orm) parsing'.

In the current paper, we examine whether LF parsing, like phrase structure parsing, involves parallel maintenance of alternatives. We argue that LF parsing and phrase structure parsing are fundamentally different in this respect. Whereas phrase structure parsing proceeds in parallel, LF parsing operates serially. In addition, we argue that LF parsing takes its input solely from the phrase structure and not from the context in conducting its initial analysis. Of course, contextual information may ultimately be used to reach an interpretation, but we argue that this is a second stage process in the construction of an LF.

In section 2, we review evidence that phrase structure parsing occurs in parallel (Boland, 1997; MacDonald et al., 1994; Trueswell, 1996). Furthermore, the preponderance of evidence suggests that phrase structure parsing makes heavy use of discourse information; that top-down information is used in the selection (Boland and Blodgett, 2001; Crain and Steedman, 1985) and possibly in the generation (vanBerkum, Brown, and Hagoort, 1999; Hsu, 2006) of alternate parses. In section 3, we discuss motivations for investigating LF parsing as a phenomenon distinct from phrase structure parsing. Finally, in section 4, we present an experiment to show that LF parsing is done serially, using a novel task: the Incremental Verification Task (IVT).

2 Previous Research in phrase structure parsing

There are three key considerations that define any model of phrase structure parsing: how many alternatives are considered simultaneously, what information is used for the generation

and selection of alternatives, and how the revision process operates (Sturt, Pickering & Crocker, 2000). In this paper, we focus on the first consideration. With respect to phrase structure parsing, there are two broad classes of views in the literature. The first class maintains that parsing is a serial process; that only one alternative is considered at a time, as in a Garden Path Model and its descendants (Frazier and Fodor, 1978; Frazier and Rayner, 1982). The second class claims that processing occurs in parallel, and many alternatives are considered simultaneously. Significant evidence suggests that phrase structure parsing is done in parallel (Boland, 1997; MacDonald et al., 1994; Trueswell, 1996). In these types of models, as a sentence is parsed, multiple possible syntactic trees are built and considered at the same time, and one is selected from the choices.

2.1 Evidence for Parallel Processing

In everyday speech, many sentences are potentially ambiguous, but these ambiguities go unnoticed because we are quick to integrate thematic and contextual information in choosing the proper syntactic analysis. The sequence of categories in (2) can be parsed in two different ways.

(2) N-V-NP-PP

One possible parse attaches the PP to the NP. The second possible parse attaches the PP to the VP. The choice of lexical items strongly influences which parse seems most likely, so that hearers rarely notice the ambiguity. This is shown by the sentences in (3) and (4), which contain the same sequence of categories.

(3) I saw the man with a dog.

(4) I hit a man with my bat.

In the sentence in (3), the most natural reading is that the man has a dog, the analysis in which the PP attached to the NP. However, in (4), the most obvious interpretation is that the hitting occurred with the bat, an analysis in which the PP is attached to the VP. The string V NP PP can have either of these two possible syntactic parses, which are disambiguated (usually) by the combination of the verb and the lexical items in the PP. Therefore, the parser, at the point of encountering *with*, must be open to the possibility of encountering either of the structures represented in (3) and (4).

We must ask how the parser gets from a string of words to a final syntactic parse, especially in the face of ambiguity. The parser could proceed in two fashions. One possibility is that the parser considers alternative parses serially. At a choice point between two alternatives, the parser selects an alternative based upon a principled decision criterion (e.g., Minimal Attachment (Clifton and Ferreira, 1989; Frazier and Fodor, 1978)). If this parsing principle results in the selection of an incorrect analysis (i.e. if the NP-attached analysis is chosen for (4)), the parser must reanalyze the string to obtain the correct structure. The second possibility is that the parser considers alternatives in parallel, generating both the VP and NP attached analyses at the choice point. Then, upon encountering further, disambiguating, evidence, the parser selects the one syntactic parse compatible with this disambiguating information.

A different type of parallel model is a constraint-based model (MacDonald et al., 1994; McRae, Spivey-knowlton, and Tanenhaus, 1998; Trueswell, 1996). In constraint-based models, alternatives are held in parallel, and ranked according to constraints. Unlike classic parallel models, there is no separate selection stage, but simply the highest-ranked analysis is pursued. In this paper, we use the investigation of phrase structure parsing as a starting point into the investigation of parsing of LFs (which we will show are parsed serially), so our focus is not on the differences between parallel models. We therefore gloss over differences between various parallel models, instead focusing our attention on the serial-parallel distinction.

There is strong experimental evidence that the syntactic parser holds multiple alternatives in parallel, and makes an initial decision by selecting from among the various possible parses. Boland and Blodgett (2001) showed that although discourse information is not integrated immediately, the multiple structures made available by noun/verb homonyms (like *duck*) are held in parallel. Boland et al. tracked eye movements while participants read sentences with noun/verb homonyms, as in (5) and (6).

(5) She saw her duck and chickens near the barn

(6) She saw her duck and stumble near the barn

Duck can be used as a noun (5) or a verb (6). Boland et al. investigated the lexical bias and preceding discourse in encountering the homonym. They found increased first pass fixation durations reflecting lexical bias, and second pass reading slowdowns for incongruent discourses.

They took this result as evidence that lexical bias is used in the generation of alternatives, which are held in parallel (and ranked), but that discourse congruency is used in selecting from alternatives.

Pearlmutter and Mendelsohn (2000) provided further evidence that syntactic alternatives are considered in parallel. They presented participants with sentences that contain a temporary ambiguity. As the subject encounters the word *described*, the sentence is ambiguous between a complement clause, in (7), or a relative clause, in (8).

(7) The report that the dictator described the country was clearly false

(8) The report that the dictator described was clearly false

The complement clause alternative, (7), is the preferred analysis. If it is the case that processing of alternatives is strictly serial, than manipulations of plausibility of the second interpretation should have no effect. However, if processing occurs in parallel, then one may expect to see effects of plausibility manipulations on the dispreferred interpretation. To test this, Pearlmutter et al. tested the sentences in (9) and (10).

(9) The report that the dictator bombed the country was clearly false

(10) The report that the dictator bombed was clearly false

Notice that the only manipulation is the change of verb, which makes the relative clause sentence 10) less plausible, because reports are not typical patients of *bombing*. The plausibility of the embedded sentence (9) remains unchanged.

Pearlmutter et al. found that participants' reading times were slower in the condition where the relative clause alternative is implausible, as compared to when the relative clause interpretation is plausible, even while the embedded sentence analysis is pursued. This suggests that the second interpretation is under consideration even while a more prevalent analysis is being pursued, suggesting that phrase structures are processed in parallel.

Therefore, there is strong evidence that phrase structure parsing occurs in parallel. Now, let us investigate the issue of what types of information are used in phrase structure parsing.

2.2 Evidence for Use of Discourse Information in Phrase Structure parsing

It is clear that the syntactic parser is relatively quick to integrate top-down information, because as language speakers, we do not experience garden-path effects at every potential moment of temporary ambiguity. But what processes are fed by this information? Is it the case that top-down information guides the generation of syntactic alternatives or does it only guide the selection from multiple alternatives?

We have reviewed evidence the phrase structure parser considers alternatives in parallel. There is also strong evidence that the syntactic parser makes use of discourse information in determining an initial analysis. A parallel model contains both generation and selection stages to obtain an initial analysis, and we now review evidence that contextual information is utilized in both of these stages. The selection process utilizes discourse information (Crain and Steedman, 1985), such as thematic information, in the selection of alternatives. Furthermore, previous studies have shown that generation of candidate parses is not constrained by grammaticality, nor does an adult generate all syntactic possibilities unless contextually supported (Hsu, 2006). Hsu finds cases where Chinese-speaking adults fail to generate grammatical alternatives, suggesting that the generation of alternatives is not completely exhaustive. These findings would be completely unexpected if the parser only had access to syntactic information during the first-pass parse, as in a Garden Path Model.

It is clear that top-down information is used in the selection of alternatives (Crain and Steedman, 1985). Garden path effects arise in sentences like (11) because *raced* is ambiguous between the passive form of the verb and the past tense.

(11) The horse *raced* past the barn *fell*

Crain & Steedman showed that garden path effects are alleviated by manipulating the plausibility of the passive construction by changing the subject NP.

(12) Children taught by the Berlitz method pass the test

(13) Teachers taught by the Berlitz method pass the test

Although there is no syntactic difference between the sentences in (12) and (13), children are more likely to be taught by teachers (passive), than to teach teachers (past). Participants demon-

strated increased garden path effects in (13), as compared to (12). Because the children are correlated with a preference for the passive interpretation of the verb *to teach*, no garden path effect occurs. This suggests that semantic information does play a role in the selection of alternatives.

There is also evidence that contextual information can affect generation of alternatives. van-Berkum, Brown, and Hagoort (1999) showed that discourse context influences syntactic analysis. They found that context can induce a contextual garden path, which is indexed by a P600 in an ERP. vanBurkum et al. looked at sentences in Dutch that are (temporally) ambiguous between a complement clause and a relative clause. (14) is a complement clause, used most felicitously in contexts where only a single referent object of the matrix verb is available. (15) is a relative clause, used most frequently in cases with multiple potential referents.

(14) De reus waarschuwde de fee dat ze niet moest vallen
The giant warned the fairy that she should not fall

(15) De reus waarschuwde de fee dat zich had vastgeklampt niet moest vallen
The giant warned the fairy that itself had clung not to fall
'The giant warned the fairy that had clung to the giant to not fall'

vanBurkum et al. found that a P600 occurred in the reading of these sentences when the preceding context (either one or two referent) conflicted with preferred context of the sentence type. These results suggest that a discourse context led participants very early on to pursue an appropriate syntactic analysis, which was reanalyzed only after an incompatible syntactic continuation. These results support a model where contextual information is critical in the generation of syntactic alternatives.

Hsu (2006) found further evidence that contextual information influences the process of generating alternative parses. First, she found that a context can lead the parser to pursue an ungrammatical analysis, suggesting that the process by which alternative parses are generated has access to more than simply syntactic information. Secondly, she found that certain syntactic alternatives are not considered unless contextual factors push in favor of such an analysis, suggesting too, that top-down information is required in the generation of alternatives.

As a test case, Hsu tested relative clauses in Chinese. Relative clauses in Chinese have clause-final relative pronouns. This property makes it such that the relative clause marker isn't reached until the end of the relative clause, as in (16), unlike in English (17).

(16) Zangsan feichang xihuan de **na-bu** chezi
Zangsan very like DE **that-CL(car)** car
'The car that Zangsan likes very much'

(17) The boy **that** ate the hot dog left.

The clause-finality of the relative clause marker creates garden path effects, with the verb initially treated as part of the matrix clause and not part of the relative clause. However, mismatched classifiers are potential cues that the upcoming clause is a relative clause, and not a complement of the verb, as in (18) (Yoshida and Sano, 2002).

(18) na-**ben** laoshi xihuan de xiaoshuo
that-**CL(book)** teacher like DE novel
'the novel that the teacher likes'

Hsu found that Chinese speakers do not use the cue provided by the mismatched classifier. That is, when presented with a sentence with a mismatch between a classifier and the immediately following noun, (as in (18)), participants exhibited a slowdown at the head of the relative clause, as compared to non-relative clause controls, demonstrating garden path effects. This suggests that, temporarily, speakers were entertaining an analysis which consisted of a SVO sentence with a mismatched classifier. Therefore, even though there is a cue available as to an upcoming relative clause, speakers do not use this information in their online parsing to predict an upcoming relative clause. Interestingly, this suggests that Chinese speakers are initially pursuing an ungrammatical analysis, one containing a mismatched classifier, over predicting an upcoming relative clause.

However, Hsu found that Chinese adults do use this cue in a relevant context (as in van-Berkum et al. (1999)). Chinese speakers showed show no slowdown at the head of the relative clause in a context in which there are two potential referents for the head of the relative clause, suggesting that adults made use of the classifier mismatch cue. Therefore, when speakers are in a context with multiple referents, they are able to predict an upcoming relative clause when a mismatched classifier is encountered, showing that the context permits generation alternative parses that were not otherwise available. This suggests that generation of alternative phrase

structures cannot be guided by the grammar alone, but in combination with access to contextual information.

Together, these results suggest that top-down information is used in both the generation and selection of syntactic alternatives.

2.3 Review

We have reviewed experimental evidence that the phrase structure parser is parallel and makes use of top-down information. Now, we turn to examining the resolution of ambiguities that do not correspond to alternative phrase structural analyses.

3 On comparing Phrase Structure parsing and LF parsing

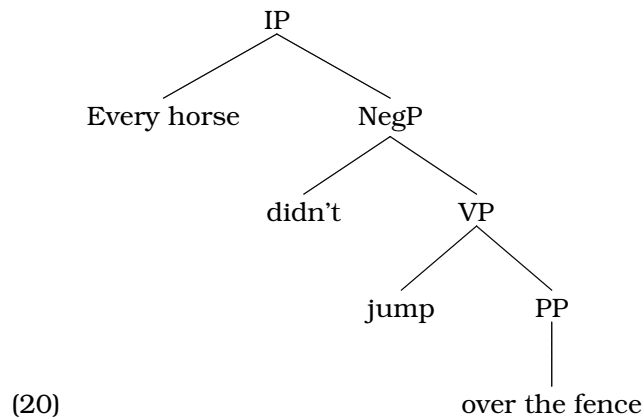
Just as we have been interested in how a speaker gets from a string of words to hierarchical structure, we must be interested in how one gets from a syntactic tree to the level of meaning; namely LF parsing. This issue has only been occasionally investigated (Kurtzman and MacDonald, 1993; Tunstall, 1998; Anderson, 2003; Carlson, Dickey, and Kennedy, 2005; Carlson, 2002). Previous results inform us about the nature of specific lexical items, but not the processes that derive from the specific nature of LF parsing. In this section, we will support our motivation for the investigation of LF parsing as a distinctly different phenomenon from phrase structure parsing.

First, we will review some offline measures for the LF and phrase structure ambiguity, then turn to evidence concerning serial versus parallel processing.

3.1 LF Ambiguity

In reviewing the phrase structure parsing literature, we saw examples where one string of words gave rise to multiple possible syntactic structures. Similarly, there are instances in which one surface tree gives rise to multiple possible LFs. One example of this is scope ambiguity. The sentence in (19) has only one phrase structure tree, as shown in (20).

(19) Every horse didn't jump over the fence



However, from the one tree shown in (20), we have two LFs, as shown in (21).

- (21) a. $\forall x[horse(x) \rightarrow \neg jump(x)]$
 b. $\neg \forall x[horse(x) \rightarrow jump(x)]$

These two LFs correspond to the two different interpretations. (19) can mean either (21a): *none of the horses jumped* (the isomorphic, or surface scope, interpretation) or (21b): *not every horse jumped* (the non-isomorphic, or inverse scope, interpretation). The precise details of the alternative LFs are not critical for this study, but only that there are two LFs corresponding to a single surface constituent structure and that the difference between them lies in the relative height of negation and the subject quantifier phrase in the semantic representation.

With respect to resolving LF ambiguity, it appears that adults arrive at an interpretation quite rapidly. Musolino, Crain & Thornton (2000) showed that when the inverse scope interpretation of a scopally ambiguous sentence is true, adults easily access it. They conducted a study using a Truth Value Judgment Task (TVJT), in which a story is acted out with characters, and a puppet attempts to describe what happened in the story. The participant's task is to judge whether the puppet said something right or wrong about what happened in the story. An example story from this experiment follows.

This is a story about three horses. The horses want to have a jumping contest to decide who will be the best jumper. They see a barn, and want to see who can jump over it the highest, but a barn is a little high, and that could be dangerous. So, they decide jumping over the fence would be a good test. The first two horses succeed, but

the third horse trips and hurts his ankle, and doesn't make it over the fence, as is depicted in Figure (1).



Figure 1: Picture of final scene, from Musolino, Crain, and Thornton (2000)

After this scenario, the target sentence is uttered, as in (22).

(22) Every horse didn't jump over the fence

In the scenario, the inverse scope interpretation, *not all of the horses jumped over the fence* is true because one horse failed to make it over the fence. The surface scope interpretation, *all of the horses failed to jump over the fence* is false, because two horses made it.

Musolino et al. found 100% of the adults accept this sentence, showing they have interpreted the target sentence with the inverse scope interpretation. Furthermore, as there are multiple interpretations available, this result shows that adults are quick to resolve ambiguity in light of various choices.

3.2 Global Phrase Structure Ambiguity

In this paper, our aim is to compare phrase structure parsing with LF parsing. To do so, we use ambiguity as a test case. LF ambiguity is always global ambiguity, as the two interpretations derive from the same surface tree. However, as we have reviewed, tests of phrase structure ambiguity are not always globally ambiguous. In fact, most experiments utilize a mismatch

effect that is only available when a temporary ambiguity is resolved by the parser in a way that is incongruent with incoming information.

Therefore, we need to investigate a case of global phrase structure ambiguity, that is, a case where the sentence is ambiguous even at the conclusion of the sentence. One case of such ambiguity is attachment ambiguity, as in (23).

(23) Put the frog on the towel on the table



Figure 2: Global phrase structure ambiguity

In the scene in Figure 2, (23) could be an instruction to *take the lone frog, and place it on the towel that is on the table*. Alternatively, the utterance could be an instruction to pick up *the frog that is on the towel, and place him on the table*. This construction is a global phrase structure ambiguity, analogous to the scope ambiguity case reviewed above.

There has been some work conducted with this construction in both adults (Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy, 1995) and children (Trueswell, Sekerina, Hilland, and Logrip, 1999). Trueswell et al. (1999) investigated sentences like (23), in either a one referent context (only one frog), or a two referent context, as is shown in Figure (3). Because Trueswell et al. were investigating the effects of the number of referents on eyegaze, they were investigating a scene in which (23) is only temporarily ambiguous.

At the point of hearing *put the frog on the towel*, the sentence is ambiguous. The sentence could be an instruction to *take the lone frog, and place it on the empty towel*. Alternatively, the utterance could be a reduced relative clause, indicating the participant should pick up *the frog that is on the towel*, and wait for an instruction about where to place it. However, as *on the table* is heard, it is only this second interpretation that remains available. Notice, this is different from the globally ambiguous case we reviewed above because there is no picture available with *a towel on the table*.

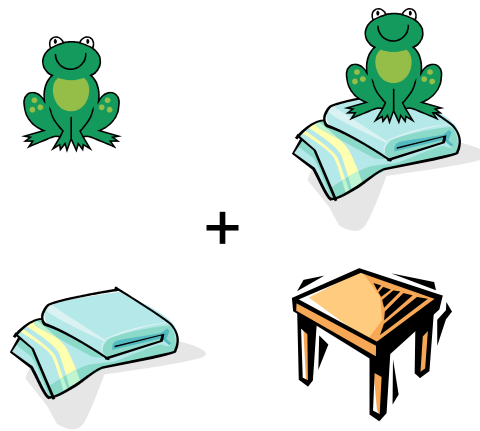


Figure 3: Example from Trueswell et al. (1999)

In an eye tracking study, Trueswell et al. found that adults were able to make use of the two referent context (and determine that there was no unique frog, and assume the PP must be a modifier). Therefore, here too, it appears that adults are able to utilize contextual information to resolve ambiguity.

3.3 Review

We are interested in investigating LF parsing, as compared to phrase structure parsing. Because LF ambiguity is always globally ambiguous, and studies of phrase structure parsing typically rely on temporary ambiguities, we require study of similar global ambiguities in both domains. Furthermore, there has been some previous work to show that with both types of ambiguities, adults are quick to integrate discourse information and settle on one interpretation. Therefore, at least at first glance, we may suspect that LF parsing works the same way as phrase structure parsing. In the next section, we review the need for a novel task to compare parsing between these two types of structures, and show that LF parsing is fundamentally different from phrase structure parsing.

4 Experiments

In the same way we investigated the nature of the syntactic parser, we want to investigate the LF parser. There are two possibilities in comparing LF and phrase structure parsing. First, it

could be that the language processor has only one way of handling any type of ambiguity, which is to generate every alternative possible, and pick from the choices. We saw that this is the case for syntactic ambiguity, and for lexical ambiguity (Boland and Cutler, 1996). For LF parsing, this would mean to generate every possible LF from a syntactic tree, and pick based on some selection criteria. The second possibility is that the nature of parsing derives from the nature of the input. That is, the architecture of different subparts of the language understanding device is likely to differ based on the information sources it uses to construct a representation.

We present an experiment to determine whether the LF parser is serial or parallel. Additionally, for the case of scope ambiguity, we suggest that the default parse is the one that matches the scope in the syntactic tree, without referencing discourse information, different from phrase structure parsing.

We have seen that with respect to phrase structure parsing, there is strong evidence for parallel processing of alternatives. Therefore, we want to ask how LF ambiguity is resolved. Does the LF parser work in serial or in parallel? Just as with phrase structure parsing, there are two hypotheses: one, the parser maintains multiple representations in parallel; that is, both are actively pursued, and could be selected from at any time. The second hypothesis is that the parser is serial, with only one alternative generated and actively pursued, requiring revision if the analysis is incorrect.

4.1 The IVT as a Method for Examining Parsing

First, we need to find a method to examine if phrase structure and LF parsing are done in parallel or serially in adults. As we have seen, the study of parallelness in syntactic ambiguity is frequently done by self-paced reading or eyetracking while reading (Hsu, 2006; Boland and Blodgett, 2001; Crain and Steedman, 1985). In these experiments, researchers look for a slowdown when incongruent syntactic information is encountered, as evidence of a violation in expectations. However, the multiple representations available in LF ambiguity derive from one syntactic representation. Therefore, it is not possible to determine, along the course of the sentence, when one grammatical interpretation is no longer available. An exception to this is Anderson (2003), who used a self-paced reading task to investigate scopally ambiguous sentences, such as (24).

(24) Every climber scaled a mountain

In these cases, the readings correspond to distributivity, where one reading has one mountain, and the second interpretation requires as many mountains as there are climbers. This allows one to look for an incompatibility between the listener's expectations about the number of mountains and the actual number of mountains (similarly in Kurtzman and MacDonald (1993) and Tunstall (1998)). However, this is a result of one of the quantifiers contributing to a distributive interpretation, and therefore, does not generalize to all types of scopal ambiguity (see Gil (1982) and Beghelli (1995) for evidence that distributivity and scope are separate phenomena). Therefore, our ability to use self-paced reading tasks for the investigation of LF ambiguity is limited. Previous experiments testing for LF ambiguity for sentences containing a quantifier and negation only look at the preferred interpretation, not the real time expectations of the participant.

Therefore, we need a task where we can determine which interpretations are being entertained. Traditionally, for syntactic ambiguity, the syntactic structure the listener expects to come up has been used as a way of determining which analyses they are pursuing. For LF ambiguity, we can look for what pieces of information the participant requires to evaluate a given interpretation. More specifically, we can look at what information the person expects to encounter to verify if a statement is true or not. For example, to verify if the sentence in (25) is true, a speaker builds a model of their expectation about houses.

(25) All houses are red

In this case, one would expect, if the statement is true, to encounter only houses that are red. Finding a house that is blue is sufficient information to declare the sentence false, without looking at any further houses. We created the Incremental Verification Task (IVT), where the information required to verify/implement the sentence is presented serially, a task which exploits speakers' abilities to build models about what a sentence means. Crucially, because we cannot examine the incremental addition of information during the sentence, we will investigate the information needed to verify the sentence.

We developed the Incremental Verification Task (IVT), in which the evaluation of the truth conditions of a sentence is carried out incrementally. This task makes both readings of an ambiguous sentence available, and asks whether adults stop at the interpretation which can be

verified/implemented first. If adults verify as soon as possible, this suggests that adults have both interpretations available simultaneously, and are able to select the proper interpretation based on truth condition. However, if adults delay in verification (can not stop at the earliest one), this suggests the first interpretation was unavailable to them.

In IVT, four pictures are hidden underneath four cups. The participant is given a target sentence, such as (26), and is told that their task is to verify if the sentence is true or false as soon as possible, i.e. with uncovering the fewest number of cups (a variant of this task is that the participant must perform an action as soon as possible, and will be reviewed below). The only rule of the game is that the cups must be uncovered from left to right. Let us demonstrate the task with an example that is unambiguous in this context (because it is pragmatically infelicitous that every zebra shares the same crown).

(26) Every zebra has a crown

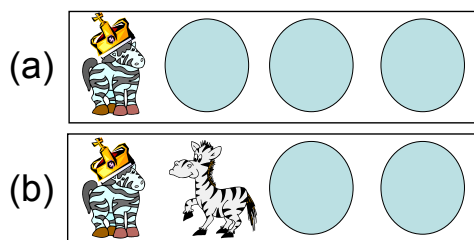


Figure 4: Incremental Verification Task: unambiguous example

To begin, the participant is told they need to verify the truth of the sentence in (26). The participant reveals the first picture, as in Figure (4a). At this point, the picture is consistent with all the zebras having crowns, but the participant can not verify the truth of the sentence. So, the participant reveals a second picture, as in Figure (4b). At this point, the participant encounters a hatless zebra, and he/she can say the sentence is *false*, without turning over any additional cups. By exploiting this incremental verification process, we can identify which interpretations adults are considering, much in the same way surprise at a continuation has been used to judge the intended structure with temporary ambiguity.

If this task is to be informative about LF parsing, then we must first verify that this task confirms previous results that phrase structure parsing is done serially. Let us now turn to the

Structural ambiguity task.

4.2 IVT: Structural Ambiguity

We have reviewed evidence that syntactic ambiguity is processed in parallel. Therefore, in an IVT where two interpretations are ultimately available, we predict that adults will stop at whatever interpretation that is available to be implemented first. That is, if both interpretations are available (as is the case if interpretations are processed in parallel), both interpretations should be available. In order to test this, we created an IVT with a PP attachment ambiguity, as in (27).

(27) Put the cat on the sofa on the rug

(27) is ambiguous between an analysis treating the PP headed by *on* as attached to NP or VP, as shown in (28).

- (28) a. [_{VP}Put [_{NP} the cat] [_{PP} on [_{NP} the sofa [_{PP} on the rug]]]]
b. [_{VP}Put [_{NP} the cat [_{PP} on the sofa]] [_{PP} on the rug]]

The NP attached PP (28a) can be paraphrased as *pick up the cat and put it on the sofa that's on the rug*, whereas the VP attached PP (28b) can be paraphrased as *pick up the cat that is on the sofa and put it on the rug*. If it is the case that syntactic alternatives are processed in parallel, and the IVT can reflect this process, then we expect adults to pick whichever interpretation for which there is sufficient information for first.

In this task, participants are told to perform the indicated action as soon as possible. We conducted an IVT experiment with two conditions. Across both conditions, the first picture is a cat, and the second cup reveals a rug. However, in one condition, the NP attached reading could be verified on the third cup, as in Figure (5). (b) reveals a sofa on a rug, so an appropriate response is to place the lone cat on the sofa.

The second condition makes the VP attached reading available on the third cup, as in Figure (6). (b) reveals a cat on a sofa, so an appropriate response is to take that cat and place it on the rug.

Notice that in both conditions, the opposite interpretation can be obtained by lifting the final cup (Figure (5c) and Figure (6c)). This way, if the participant only gets one interpretation, this

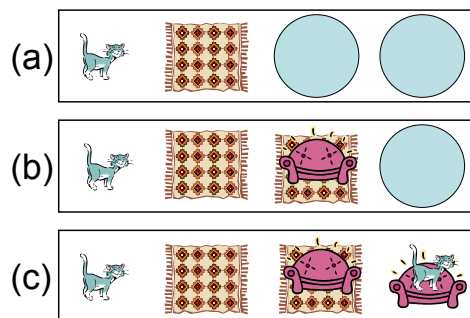


Figure 5: IVT: Syntactic Ambiguity, NP attached Condition

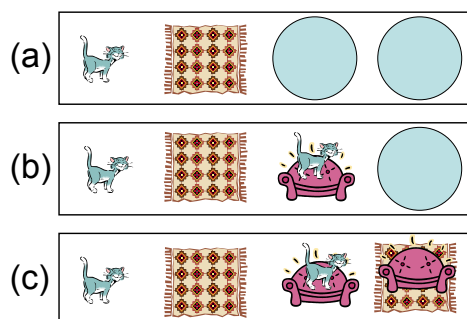


Figure 6: IVT: Syntactic Ambiguity, VP attached Condition

interpretation will be able to be acted out after the removal of the final cup. The participant's task is to perform the action stated by the experimenter.

If syntactic ambiguity is processed in parallel, then we expect the condition to be highly predictive of the reading obtained by the participant. That is, in the VP attached condition, we expect adults to pick up the cat on the sofa, but in the NP attached case, we expect participants to pick up the lone cat. However, if for some reason, the IVT as a task restricts people from considering two readings, then we expect people to stick with one interpretation across conditions. That is, the participant's interpretation would not be predicted by condition.

4.2.1 Materials and Methodology

At the start of the session, the participant is told they will be shown a series of cards, with pictures hidden underneath cups. The experimenter will utter a sentence, and participants are told that their task is to perform the action indicated, but by uncovering as few of the cups

as possible (from left to right). The cards are laminated, and designed so that each item can be revealed individually. Underneath the cups, each card contains pictures, some of which are attached with hook-and-loop tape, so that the items may be moved in order for the participants to conduct the required action. For the example in (6), both cats would be movable.

The experiment is a within-subjects design, with two conditions: the NP attached first and VP attached first. Each participant received three trials per condition, in random order. Each participant additionally received two warm-up trials, as an introduction to the task, and four filler trials interspersed among the targets. Filler trials were trials similar to the targets, but contained relative clauses that force an NP (29) or VP (30) attachment.

(29) Put the frog that's hanging onto a branch on the dresser

(30) Put the blue rug next to the chair that has a cat on it

The card for (29) is depicted in Figure 7. Order of presentation was counterbalanced. Each sentence was verbally presented by the experimenter with a prosody that did not favor either interpretation. The entire session lasted approximately 15 minutes.

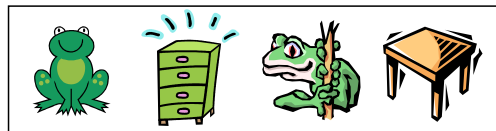


Figure 7: IVT: Syntactic Ambiguity, NP-attached Filler

4.2.2 Participants

Participants were 20 undergraduate students from the University of Maryland, and were paid for their time. All participants in the experiments described in this paper are native English speakers. No participants were excluded from analysis.

4.2.3 Results

Trials were scored on both the action, and the item the participant stopped on. If the action and cup number did not correspond to one of the interpretations, that trial was not scored. We find that condition is highly predictive of interpretation. For each condition, we calculate the

% NP attached readings. In the NP condition, we see 87% NP attached readings. In the VP attached condition, we see 21% NP attached readings, a significant difference between conditions ($t(19)=2.09$, $p < .01$). Therefore, the condition is highly predictive of the interpretation. On unambiguous controls, participants performed the proper movement, while stopping at the appropriate cup, 100% of the time.

4.2.4 Discussion

In this experiment, we found that the condition was highly predictive of interpretation. This suggests that alternative phrase structures are handled in parallel, as participants are able to obtain either interpretation depending the order of presentation of the pictures. This task produces results compatible with conclusions derived from in online reading studies. Let's turn to a test of LF ambiguity.

4.2.5 IVT: Universal Quantifier

We have confirmed that the IVT replicates the results for phrase structure parsing. Now, we want to turn to LF parsing. Because many experiments have been done offline with the interaction between *every* in the subject position and negation, these are the sentences we investigate in this paper.

Before we go over the details of the task, we must first review a peculiarity of scope ambiguity with a universal quantifier, as in (31).

(31) Every teacher can't afford a mansion

As we have reviewed, there are two interpretations to this sentence. Either *none of the teachers can afford a mansion* (surface scope), or *not all of the teachers can* (inverse scope). However, if it is the case that none can, it automatically follows that not all can. This means that there is no situation which exclusively permits the surface scope interpretation. Therefore, TVJTs must test the truth / falsity of the inverse scope interpretation. Interestingly, the IVT (although designed to investigate parallelness of parsing) remedies this problem, because the cup revealed can uniquely determine the subject's interpretation for either possible reading.

We used the IVT to test how adults verify scope ambiguities with a universal quantifier in subject position and negation, such as (32).

(32) Every dog isn't wearing a hat

The experiment is designed such that the inverse scope interpretation (the preferred interpretation for adults) can be verified (and is true) on the first cup. This is shown in Figure (8a), in which the first cup reveals a hatless dog.

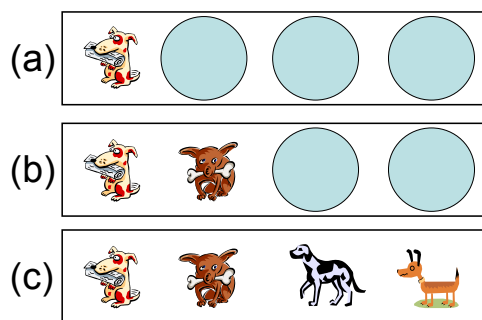


Figure 8: IVT: Universal Quantifier, Surface Scope-true

If adults hold alternative LFs in parallel, and can verify either interpretation, they will stop at the first possible interpretation, in this case, the inverse scope interpretation. Alternatively, if adults interpret the sentence according to their preferred interpretation (inverse scope interpretation in offline tasks), we would also expect the participant to stop when the inverse scope interpretation can be verified, which is at the first cup. The surface scope interpretation requires turning over more cups, and thus, is not allowed according to the rules of the game, unless the inverse scope interpretation was not under consideration. In Figure (8), this means advancing to the end (c), because the participant must verify that none of the dogs have hats. Notice that even though both interpretations have the same truth values in this situation, the interpretation can be determined by the cup on which the participant stops. This allows both interpretations to be directly tested, a feature over the TVJT.

There are trials in which the truth values of the two interpretations differ (when the surface scope interpretation is false), as shown in Figure (9). The inverse scope interpretation can still be verified on the first cup, and the surface scope interpretation can be falsified as soon as a dog is found that has a hat, as in Figure (9b). Unfortunately, due to the subset problem with the

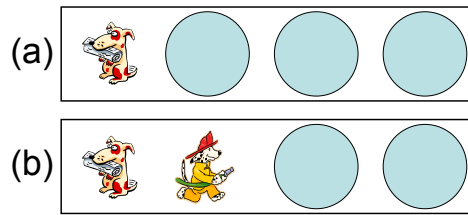


Figure 9: IVT: Universal Quantifier, Surface Scope-False

universal quantifier described above, these two conditions are the only combinations that can be tested with this task. In any scenario where the inverse scope interpretation is false, the surface scope interpretation would also be made false on the same cup.

4.2.6 Materials and Methodology

At the start of the session, the participant is told they will be shown a series of cards, with pictures hidden underneath of cups. The experimenter will utter a sentence, and the participant is told their task is to tell the experimenter if they are lying or telling the truth about what is under the cups, but by uncovering as few of the cups as possible (from left to right).

The task of the participant is to determine if the experimenter's sentence is true or false by turning over the fewest cups possible. Participants were given three trials of the target sentences, as in (32). Participants were given 3 unambiguous controls containing either *some*, as in (33) or *every*, as in (34).

(33) Some boys have a ball

(34) Every sheep has a sweater

There were both true and false control sentences. There were two types of unambiguous control sentences, those containing *every*, and those containing *some*. An appropriate response on unambiguous control trials is defined by giving a true/false response as soon as possible. For *every* trials, like (34), this means progressing to the last cup if it is the case that all of the sheep have sweaters, or stopping as soon as a sweater-less sheep is encountered. For *some* trials, like (33), an appropriate response would mean stopping as soon as one or two boys with balls are encountered. Notice however, that for (33), there is an implicature that *not all boys have*

balls. If a participant verified this implicature, this required proceeding to the final cup, and this response was scored as 'inappropriate' for the purposes of this experiment (although, for all practical purposes, this is a completely acceptable interpretation of the statement).

The order of presentation of items was counterbalanced. Each sentence was verbally presented by the experimenter with a prosody that did not favor any specific interpretation. The entire session lasted approximately 10 minutes.

4.2.7 Participants

Participants were 18 undergraduate students from the University of Maryland, College Park, and paid for their participation. No participants were excluded from analysis.

4.2.8 Results

Trials were scored on both their T/F response, and the cup the participant stopped on. If the T/F response and cup number did not correspond to the same interpretation, that trial was not scored. Across trials, the inverse scope interpretation was obtained on 23% of trials. 11/22 adults obtained only the surface scope interpretation on all three trials. Performance on target trials is significantly different than unambiguous controls ($t(17)=2.10$, $p < .01$).

Across all control trials, participants stopped on the appropriate cup on 83% of the time, and on 100% of the trials that contained *every* with no negation. Of the trials that contained *some*, participants persisted to the last cup on 33% of the trials. While this response is inappropriate for the task (as it is not stopping as soon as possible), this result does not represent a general tendency to reveal all of the cups. Only 4 adults went to the last cup on all of the *some* control items, and of these four participants, two obtained inverse scope interpretations on the target items. Therefore, the controls reveal no overall tendency to persist to the final cup.

4.2.9 Discussion

From these results, we can conclude two things. First, it appears that adults' default interpretation for scopally ambiguous sentences is the surface scope interpretation. This result confirms

previous findings from Anderson (2003) and Tunstall (1998) that surface scope is a default interpretation. These results extend this finding to a wider range of constructions containing scopal ambiguity. Second, these results show that adults, with LF ambiguity, appear fixed to their default interpretation. These results, alone, are compatible with parallel models. That is, it is possible that adults generate multiple interpretations, but then select the surface scope interpretation rapidly (before the first cup is revealed). If this analysis is correct, this would suggest that even when alternatives are considered in parallel, selection occurs too quickly to be measured by this task. However, we know from the first experiment that this analysis can not be the case. At least with phrase structure ambiguity, adults refrain from interpreting the sentence until sufficient evidence is revealed. In light of these results, it appears that LF ambiguity is not processed the same way as phrase structure ambiguity. We suggest that LF ambiguity is processed serially, with only one alternative being generated.

4.2.10 IVT: Specified Domain

One possible objection to the conclusion that LF ambiguities are parsed in serial is that the IVT restricted adults to the surface scope interpretation because the participants were lacking critical information. That is, adults, in evaluating a sentence with a universal quantifier, need to quantify over the domain, but the relevant characters were hidden. As a follow-up, we conducted an IVT where the domain is specified from the outset, as in Figure (10), with target sentences as in (35).

(35) Every cow doesn't have a hat

At the outset, all of the cows are visible. However, the participant must still overturn cups to determine what each cow possesses. Identical to the previous experiment, the inverse scope interpretation can be verified on the first trial (Figure (10a)), but verification of the surface scope interpretation is delayed (Figure (10c)).

4.2.11 Materials and Methodology

At the start of the session, the participant is told they will be shown a series of cards, with pictures hidden underneath of cups. The experimenter will utter a sentence, and the participant is told their task is to tell the experimenter if they are lying or telling the truth about what is under the cups, but by uncovering as few of the cups as possible (from left to right). After the

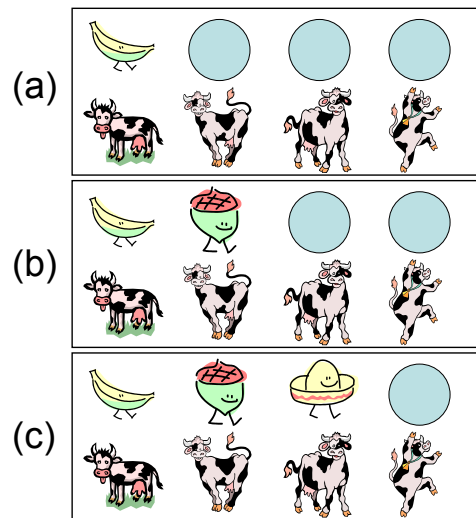


Figure 10: IVT: Universal Quantifier Specified Domain

participant has made a judgment, the experimenter presents the next card. On these cards, the animals on the index card remain visible the entire time, and are not covered by cups.

Each participant received 6 target sentences and 5 unambiguous control sentences. The unambiguous controls consisted of three trials with the quantifier *some*, as in (36), and three trials containing *every* with no negation, as in (37).

(36) Some boys have a ball

(37) Every sheep has a sweater

As in the previous experiment, there were both true and false unambiguous controls. Order of presentation was counterbalanced. Each sentence was verbally presented by the experimenter with a prosody that did not favor either interpretation. The entire session lasted approximately 15 minutes.

4.2.12 Participants

10 participants were involved in this experiment and were from the University of Maryland, College Park, and paid for their participation. No participants were excluded from analysis. One participant was replaced in the design because more than half of his/her responses did not correspond to either interpretation.

4.2.13 Results

Trials were scored on both their T/F response, and the cup they stopped on. If the T/F response and cup number did not correspond to the same interpretation, that trial was not scored. In this experiment, across target trials, 29% of the responses were consistent with the inverse scope interpretation, almost identical to IVT: Universal Quantifier. Also similar to IVT: Universal Quantifier, 50% of the participants obtained the surface scope interpretation on all trials. There was no significant difference between this experiment and the Universal Quantifier experiment where all items were covered at the outset ($t(26)=2.05$, $p > .56$).

Unambiguous controls were scored in the same manner as the previous experiment. Over all control trials, participants stopped on an appropriate cup in 74% of unambiguous controls. As in the previous experiment, the 'inappropriate' responses were entirely from the *some* trials.

From these results, we can conclude that the isomorphism observed in previous experiment is not due to a lack of information in the domain.

4.2.14 Conclusions on Seriality

In our experiments, we confirmed that phrase structure ambiguity is processed in parallel, as adults are able to alter their interpretations based on upcoming evidence. Defining parallel processing as the capability to select from multiple interpretations at any given time, we investigated whether LF ambiguity could be handled this same way. Our results from the series of IVT experiments suggest that LF parsing is handled serially, as adults appear unable access the inverse scope interpretation before verification of the surface scope interpretation.

5 Conclusions

In this paper, we investigated the nature of LF parsing. We introduced a novel task: the Incremental Verification Task, which allows us to identify analyses pursued for constructions containing global ambiguities. Using the IVT, we confirmed previous findings that multiple phrase structure representations are considered in parallel. Next, we investigated a case of LF ambiguity, sentences containing a scope ambiguity, with the IVT. We found that adults persisted in the

task to verify the surface scope interpretation, even though the inverse scope interpretation was available at an earlier point in the task. This is strong evidence that both interpretations were not simultaneously available to participants. This evidence is in line with the research done by (Kurtzman and MacDonald, 1993; Tunstall, 1998; Anderson, 2003), who showed that surface scope interpretations are the default interpretations, with inverse scope interpretations incurring an increased processing cost. Additionally, because we performed the same task with both phrase structure and LF ambiguity, this is significant evidence that the two processes by which these types of ambiguity are resolved differ. We now have strong evidence that not only is the surface scope interpretation the default interpretation selected, but this is the only interpretation generated due to the nature of LF parsing.

We conclude that LF parsing differs significantly from phrase structure parsing. While much research suggests that phrase structure parsing is processed in parallel and uses discourse information, we find that LF parsing is serial, adhering to the surface scope interpretation as a default.

Acknowledgments

This research was funded by an NSF grant to the second author (#BCS-0418309). We would like to thank all of the subjects at the University of Maryland for participating in these studies. Also, thanks to all the members of the CNL lab who have contributed comments throughout all stages of this work.

References

- C. Anderson. *The Structure and Real-Time Comprehension of Quantifier Scope Ambiguity*. PhD thesis, Northwestern University, 2003.
- F. Beghelli. *The phrase structure of quantifier scope*. PhD thesis, University of Massachusetts, 1995.
- T. Bever. *The cognitive basis for linguistic structures*. In J.R.Hayes (ed.), *Cognition and the Development of Language*, pages 227–352. Wiley, New York, NY, 1970.

- J. Boland. The relationship between syntactic and semantic processes in sentence comprehension. *Language and Cognitive Processes*, 12:423–484, 1997.
- J. Boland and A. Blodgett. Understanding the constraints on syntactic generation: Lexical bias and discourse congruency effects on eye movements. *Journal of Memory and Language*, 45: 391–411, 2001.
- J. Boland and A. Cutler. Intereaction with autonomy: Multiple output models and the inadequacy of the great divide. *Cognition*, 58:309–320, 1996.
- K. Carlson. *Parallelism and Prosody in the Processing of Ellipsis*. Routledge, New York and London, 2002.
- K. Carlson, M. Dickey, and C. Kennedy. Structural economy in the processing and representation of gapping sentences. *Syntax*, 8:208–228, 2005.
- C. Clifton and F. Ferreira. Ambiguity in context. *Language and Cognitive Processes*, 4:77–103, 1989.
- S. Crain and M. Steedman. *On not being led up the garden path: the use of context by the psychological syntax processor*. In D.R. Dowty, L. Karttunen and A.M.N. Zwicky (Eds.), *Natural Language Parsing*, page 320. Cambridge University Press, Cambridge, MA, 1985.
- L. Frazier and J. Fodor. The sausage machine: a new two-stage parsing model. *Cognition*, 6:291, 1978.
- L. Frazier and K. Rayner. Making and correcting errors during sentence comprehension: eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14: 178–210, 1982.
- D. Gil. Quantifier scope and linguistic variation. *Linguistics and Philosophy*, 5:421–472, 1982.
- C-C. Hsu. *Issues in Head-Final Relative Clauses in Chinese: Derivation, Processing and Acquisition*. PhD thesis, University of Delaware, 2006.
- J. Kimball. Seven principles of surface structure parsing in natural language. *Cognition*, 2:15–47, 1973.

- H. Kurtzman and M. MacDonald. Resolution of quantifier scope ambiguities. *Cognition*, 48: 243–279, 1993.
- M. MacDonald, N. Pearlmutter, and M. Seidenberg. The lexical nature of syntactic ambiguity. *Cognitive Psychology*, 101:676–703, 1994.
- K. McRae, M. Spivey-knowlton, and M. Tanenhaus. Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38: 283–312, 1998.
- J. Musolino, S. Crain, and R. Thornton. Navigating negative quantificational space. *Linguistics*, 38:1–32, 2000.
- N. Pearlmutter and A. Mendelsohn. Serial versus parallel sentence comprehension. *unpublished manuscript*, 2000.
- P. Sturt, M. Pickering, and M. Crocker. Search strategies in syntactic reanalysis. *Journal of Psycholinguistic Research*, 29:183–194, 2000.
- M. Tanenhaus, M. Spivey-Knowlton, K. Eberhard, and J. Sedivy. Integration of visual and linguistic information in spoken language comprehension. *Science*, 268:1632–1624, 1995.
- J. Trueswell. The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, 35:566–585, 1996.
- J. Trueswell, I. Sekerina, N. Hilland, and M. Logrip. The kindergarden-path effect: Studying online processing in young children. *Cognition*, 73:89–134, 1999.
- S. Tunstall. *The interpretation of quantifiers: Semantics and Processing*. PhD thesis, University of Massachusetts, 1998.
- J. vanBerkum, C. Brown, and P. Hagoort. Early referential context effects in sentence processing: Evidence from event-related potentials. *Journal of Memory and Language*, 41:147, 1999.
- M. Yoshida and T. Sano. Psycholinguistic studies on Japanese head internal relative clauses. In K. Hiraiwa, M. Cristina Cuervo, D. Harbour and S. Ishihara, editors, *MIT Working Papers in Linguistics*, 2002. The Proceedings of Formal Approaches to Japanese Linguistics 3.

Appendix

5.1 Sentences for IVT: Syntactic Ambiguity

Warm-up Sentences

Put the white star on the desk

Put the fly on the pink elephant

Target Sentences

Put the vase on the book on the table

Put the frog on the couch on the rug

Put the boy on the elephant on the bridge

Put the fly in the bowl on the desk

Put the star on the lamp on the dresser

Put the cookies on the chair on the blanket

Filler sentences

Put the frog that's hanging onto a branch on the dresser

Put the lamp next to the boy who's playing basketball

Put the bowl that has rice on the red couch

Put the blue rug next to the chair that has a cat on it

5.2 Sentences for IVT: Universal Quantifier

Warm up Sentence

Every zebra has a chip

Target Sentences

Every boy doesn't have a ball

Every Princess doesn't have a crown

Every frog doesn't have a fly

Every elephant doesn't have a peanut

Every monkey doesn't have a banana

Every chicken doesn't have a guitar

Filler Sentences

Some horses have a hat

Every sheep has a sweater

Some cows have a bucket

Some boys have a ball

Every squirrel has an acorn

5.3 Sentences for IVT: Specified Domain

Warm up Sentence

Every zebra has a chip

Target Sentences

Every boy doesn't have a ball

Every Princess doesn't have a crown

Every frog doesn't have a fly

Every elephant doesn't have a peanut

Every monkey doesn't have a banana

Every chicken doesn't have a guitar

Filler Sentences

Some horses have a hat

Every sheep has a sweater

Some cows have a bucket

Some boys have a ball

Every squirrel has an acorn