

ABSTRACT

Title of dissertation: THE ROLE OF VERIFICATION
STRATEGIES IN SEMANTIC
AMBIGUITY RESOLUTION IN
CHILDREN AND ADULTS

Anastasia Marie Conroy
Doctor of Philosophy, 2008

Dissertation directed by: Associate Professor Jeffrey Lidz
Department of Linguistics

This dissertation investigates the contributions of the parser and extra-linguistic information in the selection of a final interpretation of scopally ambiguous strings, integrating data from both children and adults into our understanding of language processing. Previous research has found an advantage for surface scope interpretations in adult sentence processing (Tunstall, 1998; Anderson, 2003) and in children's interpretive preferences (Musolino et al., 2000; Musolino and Lidz, 2006). In light of these findings, we investigate two central questions. One, what is the source of the advantage for surface scope interpretations in adults? Two, what factors contribute to children's ultimate adherence to surface scope interpretations? With respect to the first question, we show that the source of the advantage for surface scope interpretations cannot be described by a parsing preference, but can be described by the ease of the verification strategy utilized for the surface scope interpretation. With respect to the second question, we investigate children's interpretations of scopally ambiguous strings across a range of ages and find that while children appear fixed to

surface scope interpretations during a limited window of development, this fixation does not hold at the earliest stage of development, demonstrating a U-shaped curve of development. Additionally, we find evidence that children's interpretations do not vary as a function of task in an adult-like way, and suggest that these findings must be explained by a combination of children's developing parsers and ability to integrate discourse information. We suggest that the non-adult-like interpretations observed in children derive from an initial parser bias for inverse scope interpretations, followed by a period in which children have adult-like parsers, but lack the the ability to integrate discourse information as a means to inform the process of ambiguity resolution.

The Role of Verification Strategies in
Semantic Ambiguity Resolution
in Children and Adults

by

Anastasia Marie Conroy

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:

Associate Professor Jeffrey Lidz, Chair

Professor Norbert Hornstein

Professor Paul Pietroski

Associate Professor Amy Weinberg

Associate Professor Michael Israel

Contents

1	Introduction	1
1.1	The Advantage for Surface Scope	11
1.2	Adults and Surface Scope Interpretations	11
1.2.1	On Derivations	12
1.2.2	The Parsing Hypothesis	15
1.2.3	The Extra-Linguistic Hypothesis	18
1.3	The Isomorphism Effect in Children	19
1.3.1	Children’s Grammatical Capabilities	22
1.3.2	The Child Parsing Hypothesis	25
1.3.3	The Child Extra-Linguistic Hypothesis	27
1.4	Goals	28
2	Investigation of Timecourse of Disambiguation	37
2.1	On LF Parsing	39
2.1.1	Examining Ambiguity Resolution with the IVT	40
2.1.2	Experiment One: IVT: Universal Quantifier	43
2.1.3	Experiment Two: IVT: Number	51
2.1.4	Experiment Three: IVT: Specified Domain	55

2.1.5	Experiment Four: IVT: Reverse	60
2.1.6	Experiment Five: IVT: Written	64
2.1.7	Experiment Six: IVT: Structural Ambiguity	68
2.1.8	Experiment Seven: Speeded Forced Choice Task	75
2.2	Summary	82
3	Priming	86
3.1	Turning Adults into Children	87
3.1.1	Experiment Eight: IVT: Preceding Sentence	87
3.1.2	Experiment Nine: IVT: Prime	91
3.1.3	Intermediate Conclusions on the IVT	95
3.2	On the Nature of Priming	96
3.2.1	Experiment Ten: IVT: TVJT	101
3.2.2	Experiment Eleven: IVT: Revised Prime	107
3.2.3	Conclusions on Priming	111
4	Interpreting Scope Ambiguity	113
4.1	Revisiting the Isomorphism Effect in Children	115
4.1.1	Experiment Twelve: TVJT Reinvestigation	117
4.1.2	Experiment Thirteen: Presentation Comparison	129
4.1.3	Conclusions on Children's Abilities	134
4.2	Discourse in Ambiguity Resolution	135
4.2.1	Experiment Fourteen: Adult TVJT-Sentence Completion	135

4.2.2	Experiment Fifteen: Pilot Child TVJT-Sentence Completion	142
4.2.3	Experiment Sixteen: Child TVJT-Sentence Completion	145
4.3	Summary of the Isomorphism Effect	149
5	Testing the Parsing Hypothesis	151
5.1	Modeling Inputs to Interpretation	153
5.1.1	The Basic Model	154
5.1.2	Description of the Adult System	158
5.1.3	Experiment Seventeen: Non-biased Reading Experiment	159
5.1.4	Description of the Child's System	166
5.1.5	Conclusions	171
5.2	Assumptions about Corpora and Inverse Scope Interpretations	172
5.3	Working Memory	175
5.3.1	Experiment Eighteen: Artificial Language Word Span Experiment	179
5.3.2	Experiment Nineteen: Artificial Language and Scope Experiment	183
5.3.3	Experiment Twenty: One Word Load Experiment	186
5.3.4	Conclusions the Parsing Hypothesis	189
6	Verification	192
6.1	On Verification Procedures	193
6.2	Extended to Scope Ambiguities	197
6.3	The Existential Verification Procedure	201

6.4 Reflecting on Our Previous Findings	202
6.4.1 S-Structure vs. LF Ambiguity	203
6.4.2 On Previous Results	203
6.4.3 IVT: Number	205
6.4.4 Priming	209
6.4.5 On Children	212
6.5 Conclusions on Verification	213
7 Conclusions	215

Chapter 1

Introduction

In this dissertation, we investigate the developmental trajectory of semantic ambiguity resolution as a means of understanding how the human sentence processing mechanism extracts a semantic representation from a surface string. We take the view that investigating data from children and adults in tandem is crucial for a complete description of the mechanisms involved in interpreting a string, with each type of data illuminating the understanding of the other. A long standing finding is that children are non-adult-like in the domain of semantic ambiguity resolution. We show that this finding about children is no longer mysterious when taken in conjunction with a novel investigation of adult processing of semantically ambiguous strings. Not only does adult data illuminate our understanding of children, we also show that competing hypotheses concerning adult parsing can only be distinguished by integrating data from child parsing. Therefore, in studying children and adults together, we are able to provide a complete picture of the procedures that underlie semantic parsing. This complete picture requires the identification of the various factors that contribute to interpretation (including parser biases, discourse information, task-related effects and the way in which speakers verify interpretations). We identify the contribution of these various factors to final interpretation by viewing experimental data over a

range of age groups and over a variety of tasks. The identification of the contribution of these various factors to final interpretation allows us to construct a model of semantic parsing that can be applied to a wide range of phenomenon and describe developmental patterns in interpretation. Specifically, we find that the previously observed non-adult-like behavior in children can be described by a parser that differs from that of adults. In the domain of adult semantic ambiguity resolution, we find that preferences in ambiguity resolution can be described by appealing to adults' preferred verification strategies, a novel view that alleviates the need to attribute these preferences solely to the parser.

The study of the human sentence processing mechanism (the parser) is not a new endeavor. However, the study of the parser has focused on the processes underlying the building of surface-structure representations (Bever, 1970; Kimball, 1973; Frazier and Fodor, 1978; Crain and Steedman, 1985; Clifton and Ferreira, 1989), and not the processes underlying the building of interpretations. Strings that are locally syntactically ambiguous, such as (1), are used as a probe into this process.

- (1) The rugby player undressed on the try line was injured

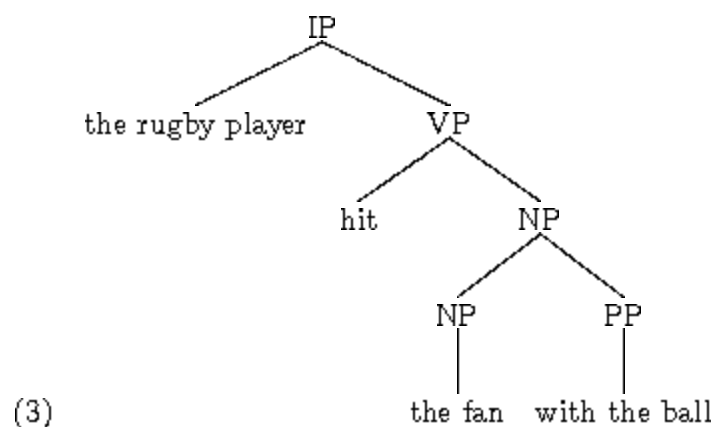
At the point of hearing *the try line*, the string is ambiguous. *Undressed* could either be a simple past tense verb, or head a reduced relative clause (which the listener discovers is the correct analysis upon encountering *was injured*). Listeners experience difficulty at encountering *was injured* (Bever, 1970), suggesting that they pursued the past tense analysis of *undressed*, the most frequent analysis of the lexical item. This observation suggests that the parser initially selects an interpretation before having the information required to conclusively determine the correct analysis. This finding initiated a line of research into what types of information are used by the parser, both in the generation and selection of alternative surface-structure analyses, and along what timecourse this information is used. Determining which analyses are under

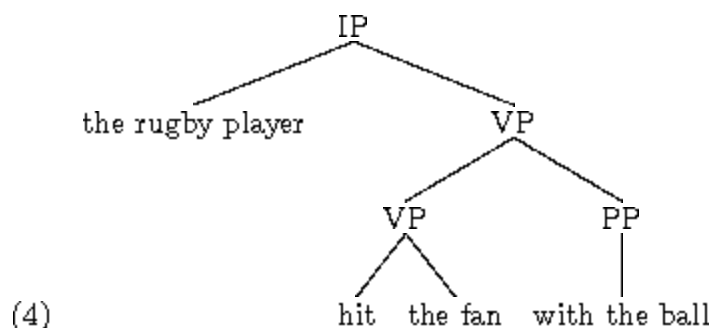
consideration and what information is utilized choosing the appropriate analysis is crucial for understanding how the parser is structured, and has been the focus of the past 30 years of research on the building of surface-structure representations (Sturt et al., 2000).

The intense study of the construction of s(urface)-structure representations has viewed the assigning of an interpretation as a trivial outcome of the process that results in an s-structure representation. This view has obscured the observation that building an interpretation may consist of parsing procedures just as complicated as building an s-structure representation. That is, even though strings containing s-structure ambiguities result in two possible interpretations, it should not be concluded that the steps involved in building these representations fully represent the complexity of the steps involved in building semantic interpretations. Let us go through an example. S-structure ambiguities can be globally ambiguous (that is, have multiple interpretations at the conclusion of the utterance, different from the example in (1)), as in the case of PP attachment ambiguity, as shown in (2).

(2) The rugby player hit the fan with the ball

In (2), *with the ball* could modify either *the fan* (as in (3)) or *hit* (as in (4)), creating an s-structure ambiguity.

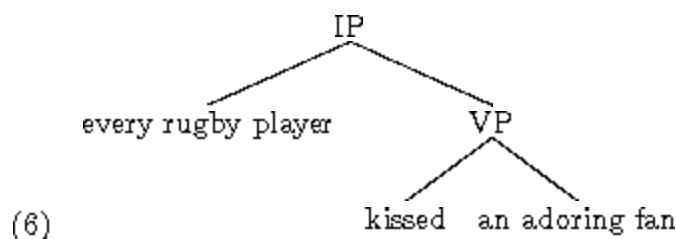




This globally ambiguous string has two interpretations, but each of these interpretations corresponds to a unique s-structure representation. From previous research on s-structure ambiguity resolution, one may conclude that there exists a one-to-one correspondence between the s-structure and semantic representations, as sentences like (2) have one interpretation per s-structure representation. However, interpretations need not be in one-to-one correspondence to s-structure representations. For example, a sentence like (5) has one s-structure representation (as shown in (6)), but yields two interpretations.¹

(5) Every rugby player kissed an adoring fan

¹Here, we are assuming a GB-style framework in which semantic ambiguity derives from covert movement, presupposing two distinct levels of representation. This is not the case in frameworks where the syntax and semantics are more closely aligned, as in Categorical Grammar (Steedman, 2001). The author considers the proper theoretical analysis to be an open question. If it is the case that only one abstract representation is constructed (playing the role of both s-structure and logical form in GB theories), then one would expect there to be no difference in the parsing procedures utilized for resolving s-structure and semantic ambiguities. Alternatively, if it is the case that there exist differences in how adults resolve s-structure and semantic ambiguities, then we may want this difference to exist in our representational theory. Although we cannot conclusively determine with this research which is the case, it is the author's hope that data from processing may differentiate between these two types of frameworks. The first step in determining the answer to this question is to investigate the existence of processing differences for s-structure and semantic ambiguities. Therefore, in this work, we will utilize a framework that makes this representational cut, although this may ultimately not be the proper representational theory.



(5) can mean either: *there is one fan that all rugby players kissed* or *each rugby player kissed his own fan*. Therefore, it cannot be the case that building an s-structure representation is sufficient to arrive at an interpretation of a string. A speaker must also build a semantic representation. Previous research on sentence processing is silent with respect to the issue of constructing a semantic representation, because only strings with one interpretation per s-structure representation have been investigated (exceptions to this include Tunstall (1998) and Anderson (2003)). However, a complete model of the parser necessitates understanding all stages of sentence processing, including the information required by the parser for the generation and selection of semantic representations.

In this dissertation, we open investigation into the study of how the parser obtains a semantic interpretation. We center our investigation on the process of coordinating a semantic representation and a phonological string, to ask about the nature of this process with respect to the parser. In an analog to the investigations that have taken place in the s-structure parsing literature, we want to determine what types of information are utilized in the initial selection of a semantic analysis. Strings that permit multiple interpretations, such as scopally ambiguous sentences, are an ideal area of investigation, because identifying how the parser selects from multiple available analyses, such as viewing the relative advantage of one interpretation over another available interpretation of the string, gives insight into the procedures utilized by the parser. From data concerning the processing of strings like (1), researchers were able to infer that the parser takes frequency information into account in the initial

selection of an s-structure representation. Similarly, we can ask what information is utilized in the initial selection of an interpretation. In this dissertation, we will center our focus on one type of semantic ambiguity, the scopal ambiguity that occurs when a string contains a quantifier that can be interpreted in a variety of positions. This type of ambiguity will lend insight into the process of semantic ambiguity resolution more generally. As a case in point, consider a string such as (7), which contains a subject universal quantifier and negation.

(7) Every rugby player can't change the rules to suit him²

(7) is ambiguous and can mean either: *none of the players can change the rules* (the surface scope interpretation) or *not all of the players can change the rules* (the inverse scope interpretation), as noted by Jackendoff (1972) and Horn (1989), among others. The domain of semantic ambiguity resolution is in its infancy, but two findings have emerged. First, it has been shown that adults initially assign surface scope interpretations to scopally ambiguous strings (containing two quantifiers), even in cases where the adults ultimately revise to obtain inverse scope interpretations (Tunstall, 1998; Anderson, 2003). Second, preschool-aged children appear to prefer surface scope interpretations of scopally ambiguous strings, even in situations where adults obtain solely inverse scope interpretations (Musolino et al., 2000; Musolino and Lidz, 2006). Together, these findings suggest that at some level, surface scope interpretations are easier to access than inverse scope interpretations, and it has been claimed that this ease is the result of a parser preference specific to the resolution of scope ambiguity. In the investigation of how the parser operates, these findings raise two questions that are the focus of investigation for this dissertation. First, what is the source of the advantage for surface scope interpretations for adults? Second, what factors account for children's non-adult-like adherence to surface scope interpretations? We

²lists.nu.ac.za/pipermail/sabirdnet/2007-August/012798.html

pursue the hypotheses that these two findings are related, in an attempt to elaborate a unified view of the parser.

With respect to both children and adults, there exist two hypotheses regarding the source of the advantage for surface scope interpretations. It may be the case that the advantage derives from decisions made by the parser, as a result of the structure of the sentence processing mechanism. Alternatively, the advantage for surface scope interpretations may be the result of preferences external to the parser. We will outline these hypotheses in detail for both age groups below.

With respect to the advantage for surface scope interpretations in adults, we consider two hypotheses: the parsing hypothesis and the extra-linguistic hypothesis. First, it may be the case that surface scope interpretations are preferred by the parser (Lidz and Musolino, 2002; Anderson, 2003). This hypothesis posits that the parser either generates only one interpretation of scopally ambiguous sentences, or that the parser generates multiple interpretations, and selects the surface scope interpretation as the most likely analysis. Second, it may be the case that the observed preference for surface scope interpretations derives from extra-linguistic factors, external to the parser. For example, if one interpretation is easier to verify than the second interpretation, then it may be preferentially selected. This hypothesis posits that the parser generates multiple interpretations, and that these interpretations carry no inherent bias associated with them. Information that is external to parsing decisions would then select from among these alternatives.

Recall that, at the most general level, we are interested in determining whether there exists an independent level of semantic parsing. Determining whether the advantage for surface scope interpretations is a result of the way the parser operates is a first step into determining the nature of semantic parsing. This determination

is a key component to identifying whether the parser treats s-structure and semantics as independent levels of representation. If we discover that there exist parsing preferences specific to semantic ambiguity resolution, then we can ask whether these preferences differ in kind from the parsing procedures that guide s-structure parsing.

Unfortunately, it is impossible to peer inside a speaker’s head and identify the processes underlying the parsing of a string. Comprehending a sentence is a psychologically complex task; not only requiring the identification of an interpretation for the sentence, but also utilizing working memory resources and verifying an interpretation against the information available in the real world. Therefore, to identify whether an observed effect is a result of parsing preferences, we must investigate the contribution of working memory and verification strategies in the sentence comprehension process. Only once the relative contributions of these factors have been identified can we make inferences about the contribution of the parser to interpretation. In this work, we investigate these components together as a way of determining the mechanisms that must be attributed to the building of interpretations from strings.

With respect to the advantage for surface scope interpretations in children, Musolino and Lidz (2006) showed that children’s deficit in accessing inverse scope interpretations is not a result of a non-adult-like grammar. With the aid of rich discourse information, children can access inverse scope interpretations, suggesting that the representations that produce these interpretations are not absent from their grammars. This finding eliminates diverging grammars as a source of the difference between children and adults, but leaves undetermined the source of this difference. Similar to the adults, there are two hypotheses regarding the source of children’s non-adult-like behavior: the parsing hypothesis and the extra-linguistic hypothesis. First, it may be the case that children have immature parsers (Lidz and Musolino, 2002; Musolino and

Lidz, 2005). Second, it may be the case that children’s adherence to surface scope interpretations are a result of immature discourse integration abilities (Gualmini, 2004; Musolino and Lidz, 2006). Determining the source of the children’s adherence to surface scope interpretations is important not only for understanding how child sentence processing occurs, but to understand the parser more generally. Child sentence processors develop into adult processors of language, so information about developmental trajectory is crucial to identifying how the parser operates. Therefore, the starting point of language processing must be taken into consideration in the development of a complete picture of the parser. In this dissertation, we consider not only ‘parser-external’ factors such as working memory and verification strategies in adults for insight into the parser, but we also consider the developmental trajectory of children’s interpretations, as an indication of the structure of their parsers, so that we may make inferences about the procedures that must be necessarily attributed to the sentence processing mechanism.

Before delving into the details, let us outline our assumptions and define our terms. With respect to the representations involved, we will be assuming a Generative Grammar semantics, as in Heim and Kratzer (1998). That is, an s-structure representation is constructed based on the surface string, and from this s-structure tree, there exists another syntactic level of representation entitled ‘Logical Form’, or ‘LF’. This Logical Form is interpreted by the semantic component, and the resulting representation will be referred to as ‘logical form’. We will use the word ‘interpretation’ to mean a logical form. We do not intend for these divisions to necessarily make any claims about processing, but this is the backdrop we will assume in service of the broader attempt to determine whether these representational levels have psychological reality. The term ‘inverse scope’ refers to a logical form in which the quantifier that is in a c-commanding position in the s-structure takes narrow scope. The term makes no

psychological claims concerning movement or inversion.

Now let us outline our assumptions concerning the parser. The parser is the culmination of processes responsible for sentence comprehension. For example, the parser can make predictions about upcoming information, such as in the case of garden path sentences, or it can utilize frequency information to select an analysis. In this dissertation, we will not focus on the internal structure of the parser. That is, there is much debate surrounding how many alternatives are generated, and how alternatives are selected (see Pickering and vanGompel, 2006, for a review). We will focus on the output of the initial stages of the parser. Specifically, in this work, we investigate the role of biases in the parser (that is, differential generation or selection of an alternative), as compared to other types of information sources external to the parser. We will refer to the process of building an s-structure as ‘s-structure parsing’. We will refer to the process of building an LF as ‘L(ogical) F(orm) parsing’. Note, however, that there is a one-to-one correspondence between LFs and logical forms, so therefore, building an LF is identical to the construction of a logical form. Therefore, the term ‘LF parsing’ actually encapsulates both of these processes. We will use the term ‘extra-linguistic information’ to include any type of information that is external to the parser. This may include the goals and desires of the individuals involved in the speech act, the events that lead up to an utterance and memory constraints. This type of information can be divided into two parts: situation and discourse. Here, we will use the term ‘situation’ to refer to any information available in the visual scene. This may be information in the real world, or information available in an experimental task. We will use the term ‘discourse information’ to refer to information about the speaker’s intentions and goals and any other pragmatic information.

Now let us turn to a review of research that demonstrates an advantage for surface scope interpretations in the process of LF ambiguity resolution.

1.1 The Advantage for Surface Scope

Strings that contain a subject universal quantifier and negation, such as (8) (repeated from (7)), yield interpretive ambiguity.

- (8) Every rugby player can't change the rules to suit him

Strings containing two quantifiers, such as (9), also give rise to ambiguity, but we focus on the construction shown in (8) because of the abundance of research available on children's interpretations of these sentences.

- (9) Every rugby player loves an Australian actress

There are two central observations in the domain of the processing of scope ambiguity. One, adults (albeit in sentences with two quantifiers, as in (9)) initially adhere to the surface scope interpretation of a scopally ambiguous string (Tunstall, 1998; Anderson, 2003). Two, five-year-old children ultimately appear fixed to the surface scope interpretation of scopally ambiguous sentences (Musolino et al., 2000). At an extremely coarse level, it appears that children adhere to the interpretation that adults initially entertain, suggesting that children are resistant to abandon this preferred interpretation. Before we outline the findings of this dissertation, let us first review the previous literature from the adult and child research.

1.2 Adults and Surface Scope Interpretations

Anderson (2003) used a self-paced reading task to investigate scopally ambiguous sentences containing *every* and *a*, as shown in (10).

(10) Every climber scaled a cliff

The string in (10) has two interpretations: either *every climber scaled his own mountain*, as shown in (11) (the surface scope interpretation), or *there is one mountain, such that every climber scaled it*, as shown in (12) (the inverse scope interpretation).

(11) $\forall x[\text{cliff}(x) \rightarrow \exists y[\text{climber}(y) \vee \text{scaled}(y, x)]]$

(12) $\exists x[\text{climber}(x) \wedge \forall y[\text{cliff}(y) \rightarrow \text{scaled}(x, y)]]$

Anderson found that, even when the discourse is biased in favor of the inverse scope interpretation, adults incurred an increased processing cost for obtaining the inverse scope interpretation of strings like (10). She additionally found the same effect for sentences like (13), which is similarly ambiguous.

(13) A climber scaled every mountain

Under the assumption that this finding reflects a general preference for surface scope interpretations in the online processing of quantificationally ambiguous strings, then these findings suggest that there is an initial advantage for surface scope interpretations.

The first goal of this dissertation is to determine the source of the advantage for surface scope interpretations in adults. We begin by outlining our theoretical assumptions concerning scope ambiguity. Then, we outline two proposals regarding the advantage for surface scope interpretations in adults.

1.2.1 On Derivations

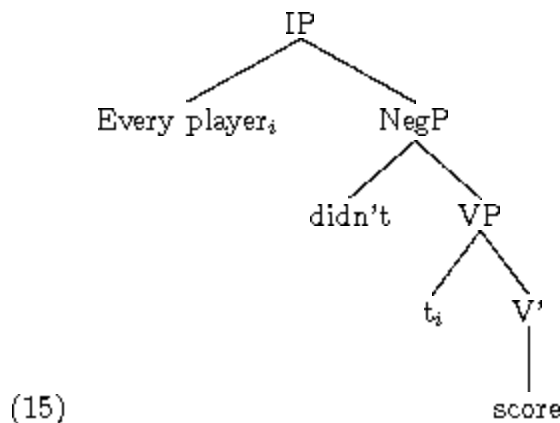
Theoretical analyses for inverse scope interpretations fall into two classes: those that posit that accessing the inverse scope interpretation requires an additional step in the syntactic derivation, as compared to the surface scope interpretation (Heim and

Kratzer, 1998; Fox, 2000), and those that do not posit any additional steps of the derivation for either interpretation (Hornstein, 1995; Steedman, 2001). In this dissertation, we investigate the contributions of the parser and discourse information in the selection of a final interpretation, and so we abstract away from the particular derivations which yield scope ambiguity. It is certainly the case that some accounts of processing scopally ambiguous strings will require the usage of theoretical accounts which posit an additional derivational step for the inverse scope interpretation, but we will show that these accounts can be translated to be based on other frameworks. Therefore, the choice of framework does not restrict us, here. We will briefly outline one account of scope ambiguity, proposed by Fox (2000), for purposes of exposition.

Fox's account of scope, like many other accounts, utilizes a level of representation called LF, a syntactic representation which feeds logical form.

(14) Every player didn't score

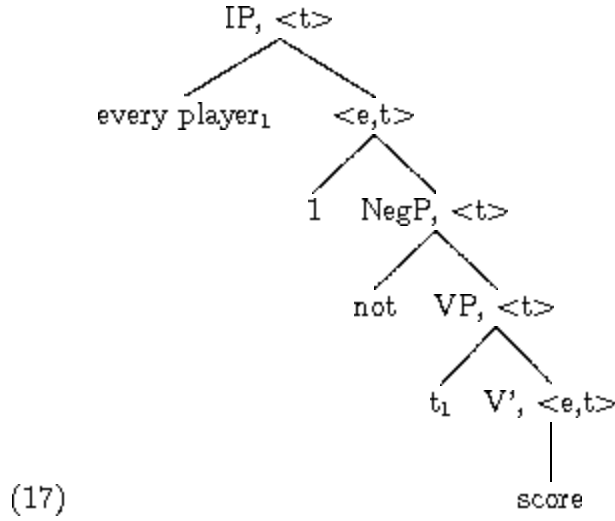
(14) has the s-structure representation in (15), where the subject has moved from its VP internal position to a case position.



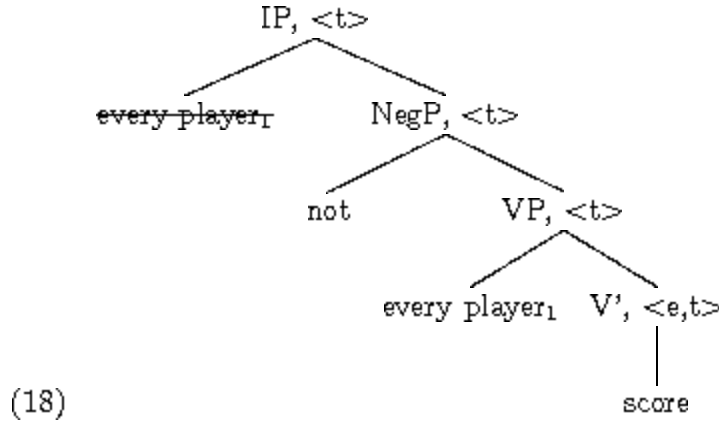
Two logical forms can be derived from (15): either (16a) (*none of the players scored*, the surface scope interpretation) or (16b) (*not every player scored*, the inverse scope interpretation).

- (16) a. $\forall x[\text{player}(x) \rightarrow \neg \text{score}(x)]$
 b. $\neg \forall x[\text{player}(x) \rightarrow \text{score}(x)]$

The LF for the surface scope interpretation of (14) can be directly obtained from (15), as shown in (17) (Heim and Kratzer, 1998).



Under Fox's account, the inverse scope interpretation of (15) is derived via an operation called 'Quantifier Lowering'. In this case, *every player* lowers to be interpreted in its VP-internal base position, as in (18).



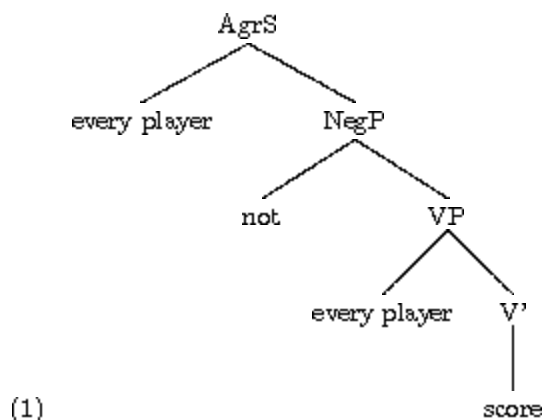
In this case, inverse scope interpretations require one extra step of the derivation over what is required to obtain the surface scope interpretation.³

³Not every account claims that inverse scope interpretations require additional steps in the derivation. Hornstein (1995), for independent reasons, attempts to eliminate covert operations such as

We now turn to two hypotheses regarding the source of the advantage for surface scope interpretations with respect to the selection of a final interpretation. One proposal suggests that the surface scope interpretation is a parsing default (Anderson, 2003; Tunstall, 1998; Lidz and Musolino, 2002; Musolino and Lidz, 2005). The second proposal suggests that surface scope interpretations gain an advantage due to extralinguistic factors.

1.2.2 The Parsing Hypothesis

The parsing hypothesis claims that the parser defaults to surface scope interpretations, and incurs an increased cost in accessing inverse scope interpretations. There are many ways that the initial preference for surface scope interpretations can be specified (even internal to the parsing hypothesis), but we restrict ourselves to outlining two: derivational complexity and PF/LF alignment. For the remainder of Quantifier Raising and Lowering. Instead of movements for QR reasons, copies are made for case checking reasons. Scope ambiguity arises because there are multiple copies of the subject NP available, as shown in (1). For interpretation, one can interpret either the higher or lower copy.



If the higher copy is interpreted, the surface scope interpretation obtains. Alternatively, if the lower copy is interpreted, the inverse scope interpretation obtains. However, this analysis may only hold of weak quantifiers (Diesing, 1992), so it is unclear if this analysis extends to the construction we investigate, here.

the dissertation, we will abstract away from the specific instantiation of the parsing hypothesis, but we outline the possibilities here for concreteness.

It may be the case that the parser prefers less complex derivations. Anderson proposed a principle of Processing Scope Economy, as stated in (19) (Anderson, 2003, pg. 63).

- (19) The human sentence processing mechanism prefers to compute a scope configuration with the simplest syntactic representation (or derivation). Computing a more complex configuration is possible, but incurs a processing cost.

Surface scope interpretations gain an advantage over inverse scope interpretations because their derivations are less complex, and therefore, are simpler for the parser to obtain. Assume that the derivation for the LF that yields the surface scope interpretation contains steps $\langle d_1, d_2, d_3 \dots d_n \rangle$. Recall that the LF that yields the inverse scope interpretation requires an additional step (quantifier lowering) to be entered into the derivation, d_{n+1} . The principle of Processing Scope Economy posits that the parser prefers the LF that yields the surface scope interpretation because its derivation contains fewer steps than the LF that yields the inverse scope interpretation. According to this hypothesis, the parser has a bias for surface scope interpretations, which is due to the fact that only the simplest derivations are generated. Although we will abstract away from much of the debate surrounding the generation vs. selection of alternative interpretations, we will outline our assumptions on the issue when required for exposition.

It need not be the case that the parsing hypothesis references derivations. An intuitive formulation of a parsing preference for surface scope interpretations is that the parser prefers string/LF pairs where a quantifier which is c-commanding at surface structure takes wide scope (J. Lidz, personal communication). It is possible that the

bias to interpret quantifiers in their surface structure position is a short-cut taken by the parser. Let us assume that the parser builds an s-structure representation and an interpretation simultaneously. This is a plausible assumption, in light of speakers' apparent ability to comprehend speech continuously and in real time. Consider an unambiguous string, such as (20).

(20) Every rugby player has their worst refereeing moment⁴

At the point of encountering *every rugby player*, the parser posits that the subject NP belongs in the [Spec, IP] (because this is the case position for the NP) in the surface structure representation. Presumably, the parser also posits a trace of movement in the VP-internal position. However, because most sentences are unambiguous, this VP-internal trace will rarely be required for scope interpretation (this position will be required for transmission of the theta role, but this is independent from scope assignment). Therefore, the parser can make a fairly accurate prediction that upon encountering a quantified subject NP, that NP belongs in [Spec, IP] at LF, and therefore, the quantifier will take wide scope in its logical form interpretation. In the case where the sentence contains negation or another quantifier, the parser will have already given the universal quantifier in subject position wide scope. This yields an advantage for surface scope interpretations. The cost of this short-cut is the occasional revision required to obtain inverse scope interpretations.

In order for the PF/LF alignment hypothesis to account for the preference for surface scope interpretations with the subject universal quantifier and negation, it must be the case that the subject quantifier is interpreted before negation is encountered. Otherwise, there would be no explanation for the advantage for surface scope interpretations. We should hope to find independent evidence that interpretation occurs

⁴<http://sport.guardian.co.uk/columnists/story/0,,1431959,00.html>

on this timescale.

For the remainder of the dissertation, we will abstract away from these two detailed formulations of the parsing hypothesis, and consider more generally the hypothesis that the advantage for surface scope interpretations is the result of a parsing preference. Let us now turn to the competing hypothesis with regards to the surface scope advantage.

1.2.3 The Extra-Linguistic Hypothesis

The extra-linguistic hypothesis claims that surface scope interpretations gain an advantage over inverse scope interpretations for extra-linguistic reasons, that is, for reasons external to the parser. This hypothesis posits that both interpretations are made equally available by the parser, and that extra-linguistic forces are responsible for the apparent preferred status of surface scope interpretations.⁵ As with the parsing hypothesis, the detailed formulation of the extra-linguistic hypothesis is not crucial to us here, but we will outline a possible account for concreteness. As an example formulation of this hypothesis, it may be the case that surface scope interpretations gain an advantage due to psychological factors. If one adheres to Johnson-Laird’s theory of mental models, then each time a sentence is heard, a mental model, a representation of the individuals present and events and relations between those individuals, is built (Johnson-Laird, 1975). In an attempt to represent a string like (21), it is intuitively easier to hold a model of n shirtless players than it is a model where the division between the number of shirtless and shirted players is unspecified.

⁵We are presenting the parser and extra-linguistic hypotheses as discrete hypotheses, but these hypotheses fall on a continuum. It may be the case that there is a parsing bias for surface scope interpretations, but that extra-linguistic factors act to strengthen this bias. We simplify here until further detail is required.

(21) Every rugby player wasn't wearing a shirt

In this way, the ease of model building may lend an advantage to surface scope interpretations. Alternatively, if one adheres to mental logics (Braine and O'Brien, 1998), then the logical form for the surface scope interpretation may be easier to compute than that for the inverse scope interpretation. Under an extra-linguistic hypothesis, both interpretations of a scopally ambiguous string are available, and in the absence of decisive discourse information, the easiest interpretation to model will be selected by default.

A crucial piece of information for differentiating between the parsing and extra-linguistic hypotheses is determining along what timecourse the surface scope interpretation gains an advantage. Let us now turn to a review of the findings with respect to the surface scope advantage in children.

1.3 The Isomorphism Effect in Children

Recall that we roughly outlined the motivation for investigating the advantage for surface scope interpretations in children and adults as a unary phenomenon. Given that children are limited in language processing with respect to adults, it is a likely hypothesis that children's adherence to surface scope interpretations reflects the same difficulties that adults experience in a temporal fashion. For scopally ambiguous strings containing negation and a universal quantifier in subject position, as in (22), adults prefer the inverse scope interpretation, *not every American is such that they know about Jason Taylor*.

(22) Every American doesn't know about Jason Taylor

The evidence for this preference comes from asking adults what these sentences mean (Musolino et al., 2000) and observations from informal corpus data regarding the usage of these types of sentences (Musolino and Lidz, 2006).

Musolino et al. (2000), found that five year olds, in contrast to adults, exclusively access the surface scope interpretation for sentences like (22). Musolino et al. conducted a study using a Truth Value Judgment Task (TVJT). In a TVJT, the experimenter acts out a story using characters, and a puppet (a second experimenter), attempts to describe what happened in the story by uttering an ambiguous target sentence. The child’s task is to judge whether the puppet was right or wrong about what happened in the story. The task is designed so that one interpretation of the target sentence, with respect to the scenario, is true, while the second interpretation is false. The researcher can infer, by using the child’s true/false judgment, which interpretation the child obtained for the target sentence.

An example sentence from Musolino et al. is shown in (23).

(23) Every horse didn’t jump over the fence

Children are presented with a story in which there are three horses, and ultimately, two jump over a fence, and one fails, as depicted in Figure (1.1).

In this scenario, the inverse scope interpretation of (23), *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 succeeded in jumping the fence.⁶ The assumption is that, if the

⁶In these stories, the story opens with the three horses considering jumping over a rock. This sets up the possibility that none of the horses would jump over the fence, making the false answer felicitous, as in Crain and Thornton (1998). This requirement is followed for all TVJTs discussed in this thesis.



Figure 1.1: Final scene from Musolino et al. (2000)

children can access the inverse scope interpretation, they will say ‘true’, because it is true that *not all of the horses jumped over the fence*. However, if children can only access the surface scope interpretation, they will think that (23) can only mean *none of the horses jumped over the fence*, and will say ‘false’.

Musolino et al. found that 0% of 5 year olds and 15% of 6 year olds interpreted the target sentence under the inverse scope interpretation, labelling this finding the ‘Observation of Isomorphism’ (also referred to as the ‘isomorphism effect’). The children’s responses contrasted with the adults’, who accepted the target sentence 100% of the time. Musolino and Lidz (2006) replicated this effect with children aged 5;0-5;11, who accessed the inverse scope interpretation only 15% of the time.⁷

The observation of isomorphism persists with sentences like (24), even though (24) can only be interpreted with the inverse scope interpretation for adults.

⁷It is important to note that, although high acceptance rates of the inverse scope interpretation are consistent with adults’ observed preference for inverse scope interpretations, these data cannot be taken as additional evidence for such a preference. The TVJT was designed to test the availability of readings, and is not an indicator of preferences.

(24) The detective didn't find someone

Musolino et al. tested sentences like (24) in a scenario where a detective found some boys who were hiding, but failed to find others. The inverse scope interpretation of (24) is true, because there exist *some boys the detective failed to find*. The surface scope interpretation is false, because it is not true that *the detective found no one*. Musolino et al. tested 2 groups of children. The younger group, aged 3;10-5;2 (mean 4;7), accepted the target sentence, obtaining the inverse scope interpretation, 35% of the time. The older group, aged 5;2-6;6 (mean 5;7), accepted the target sentence 65% of the time. Adult controls accepted the target sentence 100% of the time. It appears that with (24), the observation of isomorphism holds, as younger children obtain the inverse scope interpretation less readily than older children and adults.

1.3.1 Children's Grammatical Capabilities

These studies show that children refrain from interpreting scopally ambiguous strings with inverse scope interpretations, but do not conclusively determine the source of this difference in interpretive abilities from adults. In principle, these results could derive from a grammatical deficiency, in which children have not acquired the grammatical competence to generate inverse scope interpretations (Musolino et al., 2000). Alternatively, children could differ from adults in other ways, such as with respect to their parsing or pragmatic abilities (Musolino and Lidz, 2006). Multiple experiments show that the children's shortcomings are not grammatical.

Musolino and Lidz (2006) showed that children's preference for surface scope interpretations in strings with a universal quantifier and negation erodes under certain pragmatic conditions. They demonstrated that when the presence of negation was made more felicitous in the target sentence, children more readily accessed the inverse

scope interpretation. Musolino and Lidz performed a TVJT with stories similar to those in Musolino et al. (2000). However, in these modified stories, all three horses jumped over a log, but only two succeed in jumping over the fence, as depicted in Figure (1.2). An example target sentence is shown in (25).



Figure 1.2: Final scene from Musolino and Lidz (2006)

(25) Every horse jumped over the log, but every horse didn't jump over the fence

In this scenario, where the presence of negation is more felicitous, acceptance rates of the target sentence rose to 60%, significantly higher than the children in the control condition, who received target sentences like (23) and accessed the inverse scope interpretation only 15% of the time. This suggests that children's apparent inability to access inverse scope interpretations cannot be grammatical. Relatedly, Lidz et al. (2004) showed that sentences requiring the same grammatical operations, but which lack negation, are interpreted by children in an adult-like manner, suggesting that the relevant grammatical operations are present in children's grammars.

Gualmini (2004) finds a similar result; that children can obtain inverse scope interpretations in felicitous pragmatic situations. He claims that the isomorphism effect is an experimental artifact deriving from the appropriateness of the target sentences with respect to the experimental situation. It is assumed that the use of negation is only felicitous if that negation contrasts with an implicit goal in the story that did not

occur. The trials that tested the existential quantifier and negation in the Musolino et al. experiment differed in the degree to which the target sentence satisfied this felicity condition. Gualmini suggested that these infelicitous target sentences gave the illusion that children prefer surface scope interpretations. Gualmini performed a TVJT with two types of target sentences, exemplified in (26) and (27). These two types of sentences differ in felicity with respect to the same story, allowing the systematic testing of the effects of felicity on interpretation. These target sentences followed a scenario in which multiple guys hide, and the firefighter finds some, but misses others.

(26) The firefighter didn't find some guys

(27) The firefighter didn't miss some guys

The inverse scope interpretations (*there are some guys the firefighter didn't find* for (26) and *there are some guys the firefighter didn't miss* for (27)) are both true. Likewise, the surface scope interpretations are both false. However, the two target sentences differ in their felicity conditions. In order for the use of negation to be felicitous, it must be the case that, in the story, the firefighter was attempting to perform an action, but then failed to. In the case of (26), the felicity condition is clearly met, as it is the case in the story that the firefighter attempted to find all of the guys, but could not succeed due to the difficulty of the task. However, this felicity condition is not met in (27). That is, it is not the case that the firefighter's goal was to miss the guys, and that he failed. Therefore, the target sentence in (27) is less felicitous than (26) with respect to the story shown.

Gualmini tested 30 children between the ages of 4;0 and 5;7 and found that children accepted (27) only 50% of the time, consistent with the observation of isomorphism. However, children accepted (26), the more felicitous target sentence, 90% of the time, indicating that the children accessed the inverse scope interpretation at adult-like

levels. Gualmini concluded that the isomorphism effect is not due to a grammatical deficit, but a result of children's inability to fit a string and situation that are not pragmatically aligned. Gualmini showed that for quantifiers headed by *some*, it is possible to reduce the isomorphism effect by manipulating the situational features. Therefore, while it appears that children in many cases prefer surface scope interpretations, we can be assured this deficit is not grammatical.

The second goal of this dissertation is to determine the source of the observation of isomorphism in children, using strings containing negation and a universally quantified subject as a starting point of investigation. Similar to our question in the adult domain, there are two hypotheses as to the source of this effect. Either the children adhere to surface scope interpretations due to a limitation in their parsers (Musolino and Lidz, 2003), or due to limitations in integrating discourse information (Gualmini, 2004). We outline these hypotheses below.

1.3.2 The Child Parsing Hypothesis

The child parsing hypothesis holds that because surface scope interpretations are a parsing default for adults (Tunstall, 1998; Anderson, 2003), children's preferences for the surface scope interpretation reflect an inability to overcome this default (Lidz and Musolino, 2002; Musolino and Lidz, 2003, 2005), as stated in (28).

- (28) Children's isomorphic interpretations reflect differences between children and adults in the operation of the parser (Musolino and Lidz, 2003, page 278)

Similar to the adult parsing hypothesis, there are a number of ways to implement this hypothesis that we will outline here, but will abstract away from for the remainder of the dissertation. We outline the range of possibilities for completeness.

Musolino and Lidz (2005) claimed that ‘the mechanisms for inverse scope, whether they are quantifier raising, reconstruction, or the use of choice functions, are extremely difficult for children to access, positing that the act accessing of an inverse scope interpretation is difficult for children.’⁸ A second possibility (at times hinted by Musolino and Lidz, following Trueswell et al. (1999)) is that children have difficulty revising. Because the surface scope interpretation is a default, children are stuck with this default interpretation. These hypotheses suggest that children are restricted to generating only the surface scope interpretation.

Alternatively, children may be limited in their parsing capacity. Let us assume that adults can maintain multiple interpretations of a scopally ambiguous string in parallel. Children, who have limited working memories, may be able to maintain only one interpretation at a time (Grodzinsky and Reinhart, 1993). If adults arrive at a final interpretation by selecting from two alternatives, but children are unable to simultaneously weigh two alternatives, this would likely be a source of interpretative differences between children and adults.

Even though there are multiple ways to encode the child parsing hypothesis, we will treat them alike for the current discussion. If it is the case that the child parsing hypothesis is on the right track, then we will require future research to differentiate between these more detailed accounts.

⁸Of course, it cannot be the case that children are incapable of performing quantifier raising, or else they would be unable to comprehend any sentence with a quantifier, such as *I saw every dog*. A charitable interpretation of this quote is that the mechanisms that underlie inverse scope interpretations are difficult for children to process when there exists another interpretation that requires fewer operations. This will clearly need to be spelled out depending on one’s theory of the adult system.

1.3.3 The Child Extra-Linguistic Hypothesis

It may be the case that the child's preference for surface scope interpretations is not due to a parsing difference between children and adults, but due to a difference in the way children and adults integrate extra-linguistic information into selecting an interpretation. This hypothesis may be formulated in a variety of different ways, which we will outline here. Children's adherence to surface scope interpretations may result either from children's non-adult-like discourse integration (pragmatic) abilities, or strictly as a result of infelicity in experimental design. Crucially, under this hypothesis, children have the capability to generate multiple interpretations of a scopally ambiguous string, but appear non-adult-like due to differences in the process of selecting an interpretation from the candidates from adults. We will outline these two possibilities below. We will distinguish between children's ability to integrate discourse information and infelicity of experimental materials in section (4.1).

First, it may be the case that children are immature discourse agents (Thornton and Wexler, 1999; Noveck, 2001; Musolino and Lidz, 2006). In general, adults integrate discourse information into the ambiguity resolution process, easily selecting the appropriate interpretation given the discourse and situational information. With respect to the scope ambiguity case specifically, when a scopally ambiguous string is uttered, it appears that adults easily use discourse information to select inverse scope interpretations whenever they are appropriate. If we observe that children lack the ability to access inverse scope interpretations, it may be the case that children have not yet learned the proper interpretive function of discourse information, and therefore, arrive at non-adult-like interpretations. Alternatively, it may be the case that children can integrate discourse information, but are unable to use this information as a means of disambiguating ambiguous sentences. For the purposes of the current discussion, we will confound these two hypotheses.

Second, it may be the case that children's preference for surface scope interpretations is a function of the experimental scenario. Gualmini showed that in certain cases, children's preference for surface scope interpretations dissolves. He showed that children's ability to access inverse scope interpretations is modulated by the felicity of the experimental design. Therefore, he suggested that children's adherence to surface scope interpretations can be completely accounted for by children's inability to accommodate infelicitous experimental conditions. Alternatively, it may be the case that the conclusions drawn from our experimental methods rest on faulty assumptions. For example, the conclusion that children obtain only surface scope interpretations is drawn from data from TVJTs. TVJTs make the assumption that if a reading is true, then the participant will assent to this reading (Crain and Thornton, 1998). The conclusions that we have reviewed are undermined if children do not adhere to this principle. That is, children may access both interpretations, but in the task, feel no need to assent to the true reading (the inverse scope interpretation). In this case, the difference between children and adults can be reduced to factors specific to the task, and may reveal no linguistically significant findings. Again, our focus here is to first determine whether children's adherence to surface scope interpretations can be described by the parser or extra-linguistic factors. In the investigation of this question, we will need to investigate the degree to which children's interpretations are influenced by task information.

Let us now turn to a summary of our goals in this dissertation.

1.4 Goals

As we have reviewed, the mechanisms underlying how a speaker obtains a logical form have not been investigated to the same degree as the mechanisms that underlie

a speaker's accessing of an s-structure representation. Therefore, we want to discover how the processes involved in extracting an interpretation from the surface string compares to the process of constructing an s-structure representation, as a way to articulate the structure of the parser for all levels of representation. Ambiguity is frequently used as a probe into the process of parsing, and will be the avenue that we pursue here. We will investigate the process of resolving semantic ambiguity, specifically, the resolution of scope ambiguity. The investigation of this process outlines the two (related) main goals that form the focus of investigation for this dissertation: to specify the source of the advantage for surface scope interpretations generally, and to identify the nature of children's interpretive differences from adults. Both of these goals serve to illuminate the processes involved in the resolution of scope ambiguity, a specific case of semantic ambiguity. We pursue the hypothesis that ambiguity resolution in adults and children are related processes, and we will use information from both of these domains to inform our articulation of the parser. In order to properly investigate these two goals, we must determine the contributing factors in selecting an interpretation in both children and adults, while identifying the role of extra-linguistic factors in the initial selection of an interpretation.

As we have discussed, there are multiple components involved in the language comprehension process. To fully understand the process involved in parsing a string, we must identify the relative contribution of each of these components to sentence understanding. Only once we have identified the effect of these factors can we determine the effect of parsing decisions on sentence comprehension. Here, we will outline three basic components that contribute to language comprehension. This list is not meant to be a comprehensive account of all of the factors that contribute to language processing, but an attempt to outline a general intuition of factors that are relevant. In the process of comprehending a string in an experimental situation, there exist many

factors that influence the choice of an interpretation. We have divided these factors into three types: preferences deriving from the parser (implementation of parsing decisions), biases deriving from the discourse, and choices made by the speaker that are influenced by the demands of the task or situation. We will briefly outline these individual factors below, and we will detail these concepts throughout this dissertation, to use these factors as footholds into illuminating our understanding of parsing procedures.

In the case of s-structure ambiguity resolution, it may be the case that the parser makes initial decisions with respect to the choice of an interpretation, based solely on the processing of information available in the string alone. For example, the fact that adults garden path in sentences like (29) (repeated from (1)) suggests that the frequency of a syntactic analysis, given a verb, may guide initial parsing decisions.

(29) The rugby player undressed on the try line was injured

Notice that here, there are two ways that a parsing bias (for the most frequent analysis) can be captured. First, it may be the case that the parser generates only one analysis, which is the analysis that is most frequent (as in Frazier and Fodor (1978)). Alternatively, it may be the case that both alternatives are generated, and frequency information is utilized by the parser to select the most frequent analysis (as in Crain and Steedman (1985)). We will not present evidence to bear on this issue here. What is important for our purposes is that when disambiguation of a sentence occurs before disambiguating evidence is encountered (that is, the information required to be confident one has selected the correct analysis), this selection may be described by a bias deriving from the parser. The presence of such biases informs our understanding of what types of information the parser utilizes in making decisions.

A second factor that may play a role in sentence processing is the integration of discourse information. Discourse biases arise from decisions made about the likelihood of an analysis, given the string and the real world. For example, Crain and Steedman (1985) showed that discourse information affects s-structure parsing. Crain and Steedman showed that garden path effects can be modulated by manipulating the plausibility of the passive construction relative to the subject.

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

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

(30) and (31) are both ambiguous at the point of *taught*, as *taught* could be a simple past tense, or a passive verb. If it were the case that only the frequency of an analysis given the verb string played a role in the selection of an s-structure analysis, then we would predict that adults would always initially select the past tense analysis, as it is more frequent than the passive. This would predict garden path effects for both types of sentences upon encountering *by*, a continuation which is consistent only with the passive verb analysis of *taught*. However, there are plausibility differences between the two sentences. Children are more likely to be taught by teachers (passive), than to teach teachers (past). If it is the case that this information can be utilized in selecting an initial analysis, then we would expect the garden path effect to be alleviated in (30), as compared to (31). This prediction was borne out, suggesting that discourse information plays a role in the selection of an analysis. Again, this result could be accounted for either by discourse information playing a role in the generation of alternatives, or in the selection of alternatives, but we will not comment on this issue here. Importantly, we must examine the role of discourse information in the process of semantic parsing.

A third factor in sentence comprehension is the contribution of task-specific demands. As an example, let us discuss the behaviors required to participate in a

TVJT, a task that is commonly used to investigate children’s interpretations of sentences. An assumption that underlies this task is that if an interpretation of the target sentence is true in the experimental situation, then the participant will assent to the target sentence (Crain and Thornton, 1998, pg. 212). This assumption has been called the ‘Principle of Charity’. This view assumes that all interpretations of the target sentence are available, and that the participant is able to select the interpretation that is true in the situation. This process, because it relies on a conversational goal, is likely quite different from using frequency information to select an interpretation. Because the TVJT relies on assumptions concerning the integration of pragmatic principles to make inferences about the interpretation obtained by the participant, it is important to consider these task demands on interpretation. Only then can we be confident in how to interpret behavioral results with respect to parsing decisions. In sum, our goal is to detail these three factors, so that they can be considered with respect to any study of the parsing of language. This investigation need not be limited to the constructions discussed in this work, as it is no doubt the case that these will be relevant for any study of language processing.

In this dissertation, we primarily investigate scopally ambiguous strings such as (32), which are used by adults with inverse scope interpretations, but are interpreted by children with surface scope interpretations (Musolino and Lidz, 2006). Here, we will outline our findings concerning the interpretation of this string with respect to the factors we have just detailed.

(32) Every American doesn’t watch Rugby League

In adults, we find that there exists no parsing bias for either interpretation of strings like (32). Therefore, we suggest that the parser generates both possibilities, and that extra-linguistic information is responsible for the selection of an interpretation. Specifically, we find that adults select the interpretation that can be most easily ver-

ified. We investigate the role of verification strategies (which have been investigated independently in the domain of psychological reasoning) in ambiguity resolution. We show that the two interpretations of strings like (32) have different verification strategies. We suggest that the verification strategy associated with the surface scope interpretation is simpler, and thus, whenever the parser makes two interpretations equally available, the ease of verification procedure acts as a contributor to selecting an interpretation. This analysis extends the number of factors that must be considered when investigating language processing. In the case of young children, we do not find it to be the case that the parser equally makes available both possible interpretations of an ambiguous string. Instead, it appears that children’s parsers give preferential output to one interpretation. We will remain agnostic as to whether their parsers generate only one interpretation, or generate both with the parser selecting an interpretation. With respect to the integration of discourse information, we suggest that adults utilize frequency information in interpreting scopally ambiguous strings, but the contribution of frequency information can be eliminated under certain experimental situations. We suggest that children do not have full command of using discourse information to influence their choice of interpretation. In this dissertation, we identify the contribution of these components with respect to the resolution of semantic ambiguity, viewed across adults and children, in order to identify the contributions of the parser. We ultimately find that the biases reported in the resolution of scopally ambiguous strings can be attributed to extra-linguistic factors, suggesting that there may not exist different parsing strategies for s-structure and semantic interpretation.

We proceed as follows. In section 2, we investigate the timecourse in which adults disambiguate scopally ambiguous strings. We introduce a novel task, the Incremental Verification Task (IVT), to investigate the information required by adults to verify a scopally ambiguous string. We suggest that, different from strings containing

s-structure ambiguity, adults disambiguate to obtain surface scope interpretations, even before situationally disambiguating cues are present. In the IVT, adults obtain largely surface scope interpretations. We want to determine what factor causes this effect, and whether this same factor is in place more generally in children and responsible for the observation of isomorphism. Therefore, we must investigate if the adult behavior (obtaining surface scope interpretations) in this task mirrors child behavior more generally. In section 3, we turn to the investigation of priming. We find that adults, in the IVT, are aided in accessing inverse scope interpretations by priming much in the same way children are in other tasks. This suggests that the same mechanism underlies the surface scope interpretations of both children and adults. Additionally, in this section, we make use of the unique opportunities provided by the surface scope interpretations obtained in the IVT to investigate some related issues about priming. In section 4.1, we turn to a re-investigation of the interpretation of scope ambiguities in young children. Previous research tested mainly older children, and made inferences about the abilities of younger children. We find, that although 5 year olds display an ‘isomorphism effect’, younger children (4 year olds) do not. This suggests that children’s claimed preference for surface scope interpretations may not be as robust as previously thought. In section 4.2, we investigate the effect of discourse information on final interpretation in children and adults. Our results with four year olds suggests that they are ‘adult-like’ in accessing inverse scope interpretations. However, we must hesitate to take results from TVJTs as significant findings about preferences. We perform a TVJT and sentence completion task with both children and adults. We find that, while adults access interpretations that vary as a function of experimental task, children do not. This suggests that, even though young children display some features of adult-like performance, they have not fully obtained adult-like interpretive abilities. In section 5, we propose a model to account for this empirical evidence. We suggest that these findings can only be accounted

for by a system in which there is no parsing bias for surface scope interpretations in adults. This conclusion suggests that the advantage for surface scope interpretations derives from extra-linguistic factors. In section 6, we investigate verification procedures as a source of the surface scope advantage. We suggest that the verification procedures that adults use for inverse scope interpretations are more difficult than those used for surface scope interpretations, the source of the surface scope advantage. In conclusion, we suggest that the surface scope advantage can be accounted for by extra-linguistic factors in adults, and that a full description of children’s semantic ambiguity resolution abilities must appeal to a developing parser and developing discourse integration abilities.

Therefore, the contribution of this dissertation is two-fold. First, we present a detailed account of the factors that contribute to LF parsing. Second, the identification of the values for these factors allow us to infer which properties of the parser are construction-specific, and which properties of the parser change over time. We find that children have parsers that initially bias their decisions about the selection of an interpretation toward the interpretation that is most frequent in the input, but that children later develop into adults who have unbiased parsers. The observed advantage for surface scope interpretations in adults derives from the way in which adults implement verification procedures for different interpretations when the parser makes multiple interpretations available. This provides a framework for the investigation of other constructions and other types of ambiguity to outline a detailed picture of the structure of the parser.

In our investigation of adults’ and children’s abilities to attach a logical form to a string, we must first identify how this process occurs in adults. In the next section, we investigate whether, for strings containing a universal quantifier and negation, adults

show evidence of an advantage for surface scope interpretations before disambiguating information is encountered.

Chapter 2

Investigation of Timecourse of Disambiguation

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 (Frazier and Fodor, 1978; Crain and Steedman, 1985; Clifton and Ferreira, 1989; MacDonald et al., 1994; Sturt et al., 2000). A critical question in this domain concerns how phrases are built, with ambiguous strings often used as a probe into the s-structure building component of understanding. It is important to remember, however, that the construction of s-structure representations is not equivalent to the construction of meaning. An s-structure must be interpreted, which necessitates the building of a logical form. In cases of attachment ambiguities, each possible attachment leads to a single logical form and so it is easy to conflate s-structure building with logical form building. But these processes are necessarily distinct. The independence of meaning and s-structure can be seen by examining strings that have only a single s-structure analysis but still have two logical forms, such as (33).

(33) One dachshund ate every bone

In order to investigate the processes of s-structure and LF parsing, we can use ambiguity (that is, multiple available representations of a single string) as a probe. That is, by viewing relative advantages of an interpretation, we can gain insight into how the sentence processing mechanism operates.

The first goal of this dissertation is to determine the source of the surface scope advantage in adults. Therefore, we must determine when in the course of the sentence comprehension process surface scope interpretations gain an advantage. From TVJTs, we find that, in situations where the inverse scope interpretation is true, children obtain surface scope interpretations by the conclusion of the process of integrating discourse information into building a logical form. We also know that by this time, adults select inverse scope interpretations. However, findings from TVJTs are uninformative about the nature of parsing, because the judgment of the string is temporally distant from the completion of sentence comprehension. In the domain of sentence processing, we can ask whether, in adults, both interpretations are maintained in parallel until the situation disambiguates between interpretations, or if the surface scope interpretation gains an advantage before this disambiguation point.

In domains outside of LF parsing, a key point of investigation is the number of alternatives considered until a string-internal disambiguation point (Frazier and Fodor, 1978; Crain and Steedman, 1985; MacDonald et al., 1994; Trueswell, 1996; McRae et al., 1998; Grodner et al., 2005, among others). While we will abstract away from the particulars of these studies, we will outline the logic underlying this investigation. In these studies, participants are presented with a temporarily ambiguous string, either in a self-paced reading or eyetracking format. A slowdown at a disambiguation point suggests that the analysis that is consistent with the disambiguation material was unavailable. For example, an early experiment by Frazier and Rayner (1982)

studied participant’s eye movements while reading temporarily ambiguous strings, as in (34) and (35).

(34) Since Ray always jogs a mile this seems like a short distance to him

(35) Since Ray always jogs a mile seems like a short distance to him

Here, the ambiguity derives from the lexical ambiguity of *jogs*, which can be interpreted transitively (34) or intransitively. (35) requires that *jogs* be interpreted intransitively (so that a *mile* may be the subject of *seems*); however, this disambiguating information is unavailable at the point of encountering a *mile*. Frazier and Rayner found evidence of difficulty at *seems* in (35), suggesting that the parser pursued a transitive analysis of *jogs*, and the intransitive analysis was unavailable by the time that the disambiguation point was encountered.

The logic of these types of studies is straight-forward; to determine whether there exist any parsing pressures to select one interpretation before the encountering of disambiguating information. We are interested in the same question with respect to scope ambiguity. With scopally ambiguous strings, however, the disambiguating information does not come from string-internal information, but from situational information. In this section, we investigate the status of alternative interpretations of scopally ambiguous strings before the situation provides disambiguation cues.

2.1 On LF Parsing

The process that we are interested in investigating is how a speaker assigns a logical form to a string in real time; namely LF parsing. This issue has only occasionally been investigated for scopally ambiguous strings (Kurtzman and MacDonald, 1993; Tunstall, 1998; Carlson, 2002; Anderson, 2003; Carlson et al., 2005). In this section,

we show that for strings containing a universal quantifier and negation, adults pursue a surface scope interpretation before encountering disambiguating situational evidence. We find that these results are consistent across different types of scopal ambiguity, but that this early disambiguation does not occur in the case of attachment ambiguity. This finding suggests that there exist differences in the ambiguity resolution processes involved in LF parsing and s-structure parsing. In section (6), we will return to an account of these differences.

2.1.1 Examining Ambiguity Resolution with the IVT

In this section, we investigate whether adults, when encountering a scopally ambiguous string, entertain multiple alternative interpretations until the point of encountering disambiguating evidence, or whether the surface scope interpretation gains an advantage at the level of maintenance of alternatives. The study of maintenance of alternatives in s-structure ambiguity is frequently conducted with self-paced reading or eyetracking while reading (Crain and Steedman, 1985; Boland and Blodgett, 2001; Hsu, 2006). In these experiments, researchers look for a slowdown when incongruent syntactic information is encountered, as evidence of a violation in the participant's expectations. However, the multiple representations available in LF ambiguity derive from a single s-structure representation. Therefore, it is not possible to determine, along the course of the string, 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 strings, such as (36).

(36) Every climber scaled a mountain

In these cases, the different readings correspond to distributivity, where one reading requires a single 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 incompatibility can occur only when one of the quantifiers contributes 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 all types of LF ambiguity (specifically, scopal ambiguities in which distributivity does not differentiate between interpretations) is limited.

We require a task where we can determine, on a moment to moment basis, which interpretations are being entertained. For the investigation of s-structure ambiguity, the syntactic structure that the listener anticipates has been used as a way of determining which analyses the participant is pursuing. For LF ambiguity, we can look for what pieces of information the participant expects to encounter to verify whether a statement is true. For example, to verify whether the string in (37) is true, a speaker builds a model of their expectations about dachshunds.

(37) All dachshunds like to have fun ¹

If (37) is true, one would expect to encounter only *dachshunds that like to have fun*. Finding a *dachshund that doesn’t like to have fun* is sufficient information to declare the sentence false, without looking at any further dachshunds. 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 string available, and asks whether adults stop at the interpretation which can be verified first. If adults make their true-false judgment as a function of which interpretation is made true or false in the situation first, this suggests that

¹www.loveyourdog.com/dachshunds.html

adults have both interpretations available simultaneously, and are able to select the proper interpretation based on truth condition. However, if adults delay in making a true-false judgment, even when the situation contains enough information to make a judgment about the truth-falsity of an interpretation possible, this suggests the first interpretation was unavailable to them.

In the IVT, four pictures are each hidden underneath a cup. Participants are given a target sentence, such as (38), and are 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 in section 2.1.7). The only rule of the task is that the cups must be uncovered from left to right, and one at a time. Let us demonstrate the task with an example that is unambiguous in this situation (because it is pragmatically infelicitous that every zebra shares the same crown).

(38) Every zebra has a crown

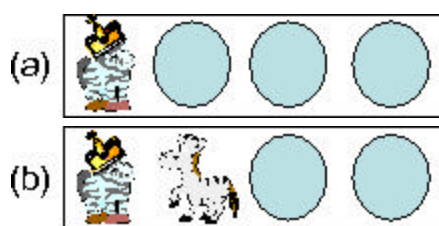


Figure 2.1: Incremental Verification Task: unambiguous example

To begin, participants are told that they need to verify the truth of the sentence in (38). The participant lifts the first cup to reveal the first picture, as in Figure (2.1a). At this point, the picture is consistent with *all the zebras having crowns*, but more evidence is needed to verify the truth of the sentence. The participant then reveals a second picture, as in Figure (2.1b). At this point, the participant encounters a hatless zebra, and he/she knows that the string 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 that a slowdown at a disambiguation point has been used to determine the structure a participant is considering in the case of temporary s-structure ambiguity.

We first want to verify whether the results found by Anderson (2003) hold for strings with a universal quantifier and negation. That is, is there an initial preference for surface scope interpretations for strings containing a subject universal quantifier and negation? If we find an initial preference for surface scope interpretations, we can determine whether this advantage derives from parsing preferences or extra-linguistic factors. We first determine whether two alternatives of a scopally ambiguous string are held in parallel at the point immediately before the situation disambiguates.

2.1.2 Experiment One: IVT: Universal Quantifier

Our interest surrounding the processing of strings with a universal quantifier and negation is two-fold. The first goal is to determine whether the advantage for surface scope interpretations holds with the construction we are investigating, and if so, to determine the factors underlying this advantage. The second goal is to determine the source of the difference between children and adults with respect to scopally ambiguous strings. As we have reviewed, children and adults arrive at different interpretations with strings containing *every* in the subject position and negation, in situations when the *not every* interpretation is true (Musolino et al., 2000). In conversation, adults use these strings with inverse scope interpretations (Horn, 1989; Musolino et al., 2000), and accordingly, arrive at inverse scope interpretations in a TVJT. However, children appear to be restricted to surface scope interpretations in TVJTs, in the absence of strong discourse information (Musolino et al., 2000; Musolino and Lidz, 2006). Therefore, to understand the source of the difference between

children and adults within this domain, we must determine the processing mechanisms involved in an adult obtaining inverse scope interpretations.

Before we go over the details of the task, we must first review a peculiarity of strings containing a subject universal quantifier and negation, as shown in (39).

(39) Every dachshund can't eat a giant bone

As we have reviewed, there are two interpretations to this string. Either *none of the dachshunds can eat a giant bone* (surface scope), or *not all of the dachshunds can* (inverse scope). However, if it is the case that *none can*, it follows by entailment that *not all can*. This means that there is no situation which exclusively permits the surface scope interpretation. Due to this subset problem, TVJTs must test the truth of the inverse scope interpretation. Interestingly, the IVT (although designed to investigate which interpretations are under consideration) remedies this problem, because as we will see, the cup revealed can uniquely determine which interpretation the participant is entertaining.

We used the IVT to test how adults verify scope ambiguities with a universal quantifier in subject position and negation, such as (40).

(40) 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 (2.2a), in which the first cup reveals a hatless dog.

The surface scope interpretation requires turning over more cups, and thus, is not allowed according to the rules of the task, unless the inverse scope interpretation is unavailable to the participant. In Figure (2.2), this means advancing to the last cup (c), because the participant must verify that *none of the dogs have hats*. Notice that



Figure 2.2: IVT: Universal Quantifier, Surface Scope-True

even though both interpretations have the same truth values in this situation, the interpretation can be uniquely determined by the cup that the participant stops on. This allows both interpretations to be directly tested, an advantage over the TVJT.

An example trial in which the surface scope interpretation is false is illustrated in Figure (2.3). The inverse scope interpretation can still be verified on the first cup.²

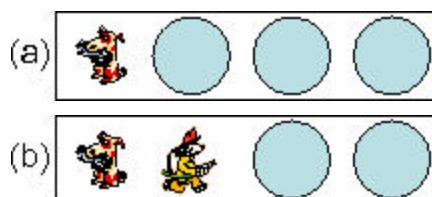


Figure 2.3: IVT: Universal Quantifier, Surface Scope-False

The surface scope interpretation can be falsified on the second cup, as shown in Figure (2.3b), when a *dog with a hat* is encountered. Therefore, obtaining the surface scope interpretation does not necessitate persisting to the final cup.

We can use this task to look at the expectations of the speaker. Because this task is carried out at the completion of the utterance of the string, it cannot reveal infor-

²It may be the case that the inverse scope interpretation carries an implication that *at least one dog has a hat*. In this case, a participant may continue to reveal the second cup. At this point, the participant, if they obtained the inverse scope interpretation, would say ‘true’. In this case, the response differs from that of the surface scope interpretation.

mation about parsing decisions made during the processing of the string. However, it does reflect sentence processing from the time the string is heard to the verification procedure. If we find that adults maintain only the surface scope interpretation, we can infer that some component of the language comprehension system lends an advantage to surface scope interpretations before the integration of discourse information. This finding alone, is a crucial step in the understanding of LF parsing, and sheds light on the biases a language comprehender brings to a TVJT.

Materials and Methodology

At the start of the session, participants are told that they will be shown a series of cards. Each card contains four pictures, each picture hidden underneath a cup. The experimenter utters a string, and the participant is told that his/her task is to tell the experimenter whether they are lying or telling the truth about what is under the cups, by uncovering as few of the cups as possible (from left to right).

The participants are given a warm-up trial and three target trials, as exemplified by (41), repeated from (40).

(41) Every dog isn't wearing a hat

Participants received both target trials that were true on the surface scope interpretation ($n=2$) and trials that were false on the surface scope interpretation ($n=2$). Participants were given 12 filler items containing either a number in object position, as in (42), a number in subject position, as in (43) or the universal quantifier in object position, as in (44).

(42) Mouse doesn't have two blocks

(43) Two zebras don't have a chip

(44) Dwarf doesn't have every fruit

There were both true and false filler items. These constructions were chosen because they display different preferences with respect to interpretation. Adults readily obtain either interpretation of (42) (Lidz and Musolino, 2002), prefer the surface scope interpretation of (43) (Musolino and Lidz, 2003), and can only obtain the surface scope interpretation of (44). Therefore, these filler items provided the participant with a variety of responses across the experiment, and contrast with the inverse scope preference observed for the target construction.

Trials were scored on both the T/F response, and the cup the participant stopped on. The true/false answer was of highest importance. If a participant gave a T/F response and persisted one cup beyond the appropriate response, then this trial would be scored by the T/F answer. For example, assume the two readings allow for the following responses: T on cup 1 or False on cup 3. A T on cup 2 would be scored as correct for the first response. However, a T on cup 3 would not be scored, because this could either be an inappropriate cup response or an incorrect T/F judgment. Therefore, in essence, there is a one-cup leniency. If the action and cup number did not correspond to one of the interpretations, that trial was not scored.

In this experiment and all subsequent IVTs, participants may be excluded for either of the two following reasons: a) failure to obtain either interpretation on more than half of the target trials,³ or b) removing cups before the completion of the target sentence.

³Recall that there are both target trials where the surface scope interpretation is true and ones where it is false. If it is the case that the participant reveals all of the cups (and says 'true'), then this is consistent with a surface scope response half of the time, but not consistent with any response when the surface scope interpretation is false (regardless of T/F response).

Target and filler items were mixed in a pseudo-random order. The order of presentation of items was counterbalanced. Two lists were created, with the second list presenting target and control items in the reverse order of the first list. Half of the participants received list one, and the other half received list two. Each string was verbally presented by the experimenter with a prosody that did not favor any specific interpretation.⁴ The entire session lasted approximately 15 minutes.

Predictions

If adults maintain both interpretations of a scopally ambiguous string up until a disambiguation point, we predict that they will be able to verify either interpretation of the scopally ambiguous target sentence. That is, we predict that the participant will stop at the first possible interpretation that can be verified, in this case, the inverse scope interpretation. This finding, would be inconsistent with the parsing hypothesis, which claims that surface scope interpretations are a parsing default, but would be consistent with the extra-linguistic hypothesis.

If adults adhere to surface scope interpretations even before the situation disambiguates, then we expect adults to persist past the first cup, in an effort to verify the surface scope interpretation. This finding is predicted by the parsing hypothesis (suggesting a parsing default for surface scope interpretations), but could also be accounted for by some formulations of the extra-linguistic hypothesis.

Participants

22 students at the University of Maryland, College Park, participated in this experiment. No participants were excluded. Participants received either monetary

⁴For a discussion of prosody and scope interpretations, see section 2.1.6.

compensation or course credit for their participation. All participants were English native speakers over the age of 18.

Results

Across all target trials, participants obtained the inverse scope interpretation 22.9% of the time. There was no difference between trials in which the surface scope interpretation was true and those in which it was false. 14/22 participants obtained surface scope interpretations on all trials. The remaining 8 participants obtained the inverse scope interpretation on 62.5% of trials. One trial was excluded from analysis because the response did not correspond to either interpretation.

On the filler trials, participants provided a response consistent with either a surface or inverse scope interpretation on 89% of the trials. This overall good performance on filler trials suggests that the participants adhere to the rules of the task.

Discussion

Participants in this task obtained largely surface scope interpretations on strings containing a universally quantified subject and negation. Even though the inverse scope interpretation could be verified first, adults refrained from verifying this interpretation, which suggests that the inverse scope interpretation was unavailable by the time disambiguating evidence was encountered. These results also suggest that there exists an initial advantage for surface scope interpretations even before disambiguating information is encountered. These results extends the findings by Anderson (2003) and Tunstall (1998) to a wider range of scopally ambiguous strings.

These results cannot be explained by a simple tendency to persist to the last cup. First, the results from unambiguous control trials show that all participants are capa-

ble of stopping on the appropriate cup, suggesting that it is not the case that adults reveal all cups before making a judgment. Second, there existed trials on which the surface scope interpretation was false, which required stopping before the last cup. Participants performed this response an equal number of times as when the surface scope interpretation was true. Therefore, there exists no overall tendency in this task to persist to the ultimate cup.

These results are surprising given two potential alternative outcomes. First, given that the inverse scope interpretation is the adult's preferred interpretation in conversation, it could have been the case that adults immediately interpreted the target sentence with the inverse scope interpretation. That is, even if surface scope interpretations obtain an initial advantage, adults could have done the following. The participant could have heard the target sentence, and then thought about what that string meant. Musolino et al. (2000) reported that when participants are asked what a string with a subject universal quantifier and negation means, they reported obtaining the inverse scope interpretation. If this is the case, one wonders why participants do not perform this interpretive process straight away. Second, adults could perform a meta-analysis at every cup. That is, after revealing the first cup, the participant could ask themselves what the target sentence means. Given that adults readily obtain the inverse scope interpretation in other situations, they should be capable of obtaining inverse scope if they revised (or checked if they needed to) at every cup. However, neither of these alternative outcomes obtained.

This result suggests that there exists an advantage for surface scope interpretations with strings containing a subject universal quantifier and negation. This finding can be explained by either of the hypotheses concerning the source of the surface scope advantage. Under the parsing hypothesis, these results are easily accounted

for by claiming that surface scope interpretations are a parsing default. Therefore, in the absence of additional discourse information, the surface scope interpretation is obtained. These results are also captured under the extra-linguistic hypothesis, if the extra-linguistic hypothesis posits an advantage for surface scope after parsing, but before the integration of situational information.

2.1.3 Experiment Two: IVT: Number

We want to determine the degree to which this early disambiguation that we have observed with strings containing a universally quantified subject and negation extends to other types of strings containing scopal ambiguity. In this section, we show that similar effects obtain in strings with negation and a number in object position, as in (45). The string in (45) has two interpretations.

(45) Jon didn't find two hearts

The surface scope interpretation (NOT > TWO) means that *it is not the case Jon found two*. The inverse scope interpretation (TWO > NOT) means that *there are two he did not find*. This type of ambiguity does not give rise to the subset problem that holds with strings containing the universal quantifier and negation. Therefore, it is possible to counterbalance which interpretation can be verified first. This construction was tested in Lidz and Musolino (2002), who found that while adults can obtain either scope, children appear restricted to the surface scope interpretation. Therefore, this construction shares an interpretive profile similar to the strings containing the subject universal quantifier and negation.

Materials and Methodology

Participants are given the same instructions as with the previous task. That is, there are a series of cards with items that are hidden under cups. The participant's task

is to verify if a string uttered by the experimenter is true or false, by revealing the fewest number of cups. Because the inverse scope interpretation requires reference to items that were not found, this version contained a modification in the design of the cards. Each card contains a wolf, who participants are told hides items, and a boy named Jon, who tries to find them. If an item is in a basket, then Jon found it, but if it is behind a leaf, then that is one that Jon did not find. Because an item is clearly marked as ‘found’ or ‘not-found’, both interpretations of the target sentence are equally felicitous.

This experiment contained two conditions: Surface-Scope first and Inverse-Scope first. In the Surface-Scope first condition, the surface scope interpretation can be verified first, as shown in Figure (2.4).



Figure 2.4: IVT: Number, Surface-Scope first

In this case, it can be verified on the second cup (Figure (2.4a)) that there are two found hearts, indicating that the target sentence is false. The alternate interpretation can be verified by revealing more cups, as shown in Figure (2.4b).

The Inverse-Scope first condition contains trials where the inverse scope interpretation is the first interpretation that can be verified, as shown in Figure (2.5).

In this case, it can be verified on the second cup (Figure (2.5a)) that there are two hidden hearts. Again, the alternate interpretation can be verified later on, as shown in Figure (2.5b).



Figure 2.5: IVT: Number, Inverse-Scope first

Each participant received 8 target trials, 4 trials from each condition. Additionally, each participant received 6 unambiguous control items, 3 with no negation, as shown in (46) and 3 with negation, but no quantifier, as shown in (47).

(46) Jon found 2 pineapples

(47) Jon didn't find any carrots

These control items serve to balance the number of true/false responses over the course of the experimental session, and to vary the cup that the participant stops on over trials.

Trials are scored according to the criteria in the previous experiment.

The conditions and unambiguous control items are mixed in a pseudo-random order. The trials are counterbalanced for order of presentation. Two lists were created, with the second list presenting target and control items in the reverse order of the first list. Half of the participants received list one, and the other half received list two. Each string was verbally presented by the experimenter with a prosody that did not favor any specific interpretation. The entire session lasted approximately 10 minutes.

Predictions

If it is the case that there exists an advantage for surface scope interpretations across constructions, and an interpretation is selected before disambiguating information is

encountered, then we would expect little difference in interpretation between conditions. That is, we predict that participants will adhere to an interpretation, regardless of the time at which that interpretation can be verified.

If it is the case that the advantage for the surface scope interpretation is specific to scope ambiguities containing a universal subject quantifier and negation, we may not expect to find a similar effect in this case. If it is the case that both interpretations are available until a situational disambiguation point, then we expect condition to be highly predictive of interpretation, because participants will be able to verify whichever condition can be verified first. That is, we would expect participants to obtain surface scope interpretations in the Surface-Scope first condition, and mostly inverse scope interpretations in the Inverse-Scope first condition.

Participants

20 undergraduate students at the University of Maryland, College Park, participated in this experiment, with one participant replaced in the design because more than half of his/her responses did not correspond to either interpretation. Participants received either monetary compensation or course credit for their participation. All participants were English native speakers over the age of 18.

Results

We are interested in determining whether condition (which reading can be verified first) is predictive of interpretation. We scored all responses on % surface scope interpretation. On the trials where the surface scope interpretation can be verified first, adults obtained the surface scope interpretation in 47% of the trials. On the trials in which the inverse scope interpretation could be verified first, adults obtained the surface scope interpretation on 40% of the trials. The two conditions are not

significantly different ($t(19) = 1.18$, $p = .24$). Across both conditions, 8/20 participants obtained only the surface scope interpretation.

Nine total target trials were excluded from analysis due to failure to correspond with either interpretation. Participants provided the correct response on 91.6% of control trials.

Discussion

In this experiment, experimental condition is a poor predictor of the interpretation obtained by the participant. Adults appear unable to use the information about which interpretation can be verified first as a cue to guide their selection of an interpretation. This result is consistent with the IVT: universal quantifier results, because adults select an interpretation before encountering disambiguating information. However, the results from this experiment differ from those of IVT: universal quantifier in that there exists no overall adherence to surface scope interpretations. What is clear from both of these experiments is that in the domain of scope ambiguity resolution, there exists some force for selecting an interpretation that is independent of situational information, although the interpretation selected may vary according to construction. We will outline some experiments to control for the results we have shown above, then turn to comparing scope ambiguity with another type of ambiguity, attachment ambiguity.

2.1.4 Experiment Three: IVT: Specified Domain

For our conclusions regarding an advantage for surface scope interpretations to hold, we must ensure that our results cannot be obtained for any task-specific reasons. Therefore, we must work to carefully control our task. In the experiments that follow,

we control for domain of quantification (section 2.1.4), which interpretation can be verified first (section 2.1.5) and prosody (section 2.1.6).

Our conclusions about the advantage for surface scope interpretations centers around the finding that adults do not make a T/F judgment at the earliest possible point. For this conclusion to hold, we must eliminate any alternative reasons for their hesitating to make a judgment. One reason adults may persist is due to uncertainty of the domain of quantification. In the previous experiment, we use the quantified noun phrase *every dog*, but participants do not know how many dogs there are. Therefore, the persistence could be due to a desire to view the entire domain of quantification before making a judgment, and not strictly due to the verification of the surface scope interpretation. Notice that, thus far, these two factors have been confounded. Persisting until the domain of quantification is revealed would produce the same results as obtaining the surface scope interpretation.

To resolve this potential confound, we conducted an IVT where the domain is specified from the outset, as shown in Figure (2.6), with target sentences as in (48).

(48) 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.⁵ Identical to the previous experiment, the inverse scope interpretation can be verified on the first cup (Figure (2.6a)), but verification of the surface scope interpretation is delayed (Figure (2.6c)).

⁵In this experiment, and in all subsequent IVT experiments, the 'cups' are actually laminated circles that are attached to the card with velcro.

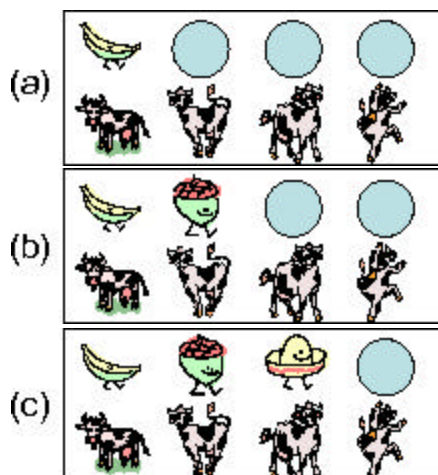


Figure 2.6: IVT: Specified Domain, target card

Materials and Methodology

The description of task to the participant and the implementation of the task remains the same as in previous IVT experiments.

Each participant received a warm-up trial, 6 target trials, 4 target trials where the surface scope interpretation is true, and 2 where the surface scope interpretation is false. Participants also received unambiguous control trials. Recall that in IVT: universal quantifier, the filler items were scopally ambiguous strings. However, because they allow multiple interpretations, the T/F answers cannot be carefully balanced across the task. To remedy this concern, we replaced the filler items from IVT: universal quantifier with 5 unambiguous control items containing a quantifier and no negation. There were two types of unambiguous control sentences: 3 items containing *some* (49) and 2 items containing *every* (50).

(49) Some horses have a hat

(50) Every sheep has a sweater

The control items serve two purposes. One, if the participant responds appropriately, these items encourage the participant to stop at a variety of cups across different trials and also creates a variety of true and false responses across the experiment. Two, these items allow us to verify that participants are appropriately participating in the task, and stopping as soon as possible.

Target items were scored with the same criteria as for IVT: universal quantifier. Unambiguous control items are similarly scored by a T/F response and cup that align with a coherent interpretation. An appropriate response on unambiguous control trials is defined by giving a true/false response as soon as possible. For *every* trials, 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, there are two appropriate responses, depending on whether the participant considers the implicatures involved. An ideal response would mean stopping as soon as one or two horses with hats are encountered. However, notice that for (49), there is an implicature that *not all horses have hats*. If a participant computed this implicature, and reported on the basis of it, this would require proceeding to the final cup.

Target and unambiguous control items were mixed in a pseudo-random order. The order of presentation of items was counterbalanced. Two lists were created, with the second list presenting target and control items in the reverse order of the first list. Half of the participants received list one, and the other half received list two. Each string was verbally presented by the experimenter with a prosody that did not favor any specific interpretation. The entire session lasted approximately 10 minutes.

Predictions

If it is the case that the apparent advantage for surface scope interpretations is due to adults attempting to view the entire domain of quantification, then we expect higher rates of inverse scope interpretations in this task, as compared to IVT: universal quantifier.

However, if the previous results are genuinely due to verification of the surface scope interpretation, then we expect little difference between the results of this experiment and the IVT: universal quantifier.

Participants

20 undergraduate students at the University of Maryland, College Park, participated in this experiment. One participant was replaced in the design because more than half of his/her responses did not correspond to either interpretation. Participants received either monetary compensation or course credit for their participation. All participants were English native speakers over the age of 18.

Results

Participants obtained the inverse scope interpretation 21.4% of the time, almost identical to IVT: universal quantifier ($t(38)=.058$, $p=.95$). 13/20 participants obtained surface scope interpretations on all trials. The remaining 7 participants obtained inverse scope interpretations on 57% of trials. 8 trials were excluded from analysis due to failure to correspond with either interpretation. Therefore, it appears the adherence to surface scope interpretations observed in the previous experiment is not due to a lack of information about the domain of quantification.

Across all control trials, participants gave an appropriate answer 100% of the time. Participants stopped on the ideal cup 74% 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 40% of the trials. While this response may be viewed as non-ideal performance on this task (as it is not stopping as soon as possible), this result does not represent a general tendency to reveal all of the cups, as adult stop appropriately on *every* control items. In this experiment, and in subsequent experiments that use this control item, there is no correlation between verification of the implicature and surface scope interpretations, suggesting this is not simply an inability to stop on the appropriate cup. Therefore, from this point on, these types of responses on *some* trials will be coded as correct.

Discussion

From these results, we can conclude that the surface scope interpretations obtained in IVT: universal quantifier are not an illusion caused by a desire to identify the domain of quantification. This control, then, strengthens our confidence in the claim that adults obtain surface scope interpretations for sentence-dependent reasons, and not for task-related reasons.

2.1.5 Experiment Four: IVT: Reverse

In IVT: universal quantifier, the inverse scope interpretation is verifiable on the first cup. We designed the experiment this way because it has been observed that in speech, adults display a preference for inverse scope interpretations (Musolino et al., 2000; Musolino and Lidz, 2006), and this experimental design made it most favorable for participants to stop on the first cup. However, we have found (and replicated) that adults persist past the first cup, leading us to conclude that adults obtain largely

surface scope interpretations in this task. Furthermore, we have determined that this persistence is not a result of a desire to verify the domain of quantification. However, we must ensure these results hold when the surface scope interpretation is the first interpretation that can be verified.

Materials and Methodology

We designed a version of the IVT where the surface scope interpretation of (51) can be verified on the first cup, using the same target sentences as IVT: specified domain, as well as the same pictures, but with the pictures appearing in a different order on the cards.

(51) Every cow doesn't have a hat

In Figure (2.7a), the first cup reveals a cow with a hat. At this point, the surface scope interpretation, *none of the cows have a hat* is false. To verify the truth of the inverse scope interpretation, *not all of the cows have a hat*, the participant must find a hatless cow. At revealing the second cup, Figure (2.7b), the participant can determine that the inverse scope interpretation is true. As in IVT: specified domain, there are

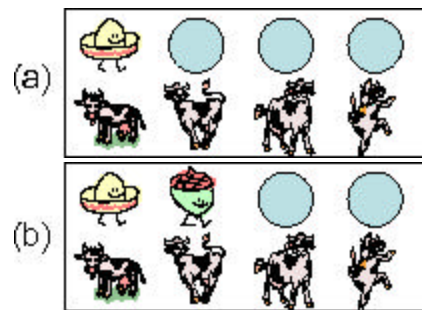


Figure 2.7: IVT: Reverse, target card

both trials where the truth conditions of the two interpretations differ ($n=4$), and trials in which they are the same ($n=2$). Crucially, though, the verification of the two interpretation always requires revealing a different number of cups.

Unambiguous controls remain the same as in IVT: specified domain. These trials, as well as target trials, are scored according to the previously mentioned criteria.

Target and unambiguous controls are mixed in a pseudo-random order. The order of presentation of items was counterbalanced. Two lists were created, with the second list presenting target and control items in the reverse order of the first list. Half of the participants received list one, and the other half received list two. Each string was verbally presented by the experimenter with a prosody that did not favor any specific interpretation. The entire session lasted approximately 10 minutes.

Predictions

Our previous results have suggested that there exists an advantage for surface scope interpretations, even before disambiguating information is encountered. This conclusion makes the prediction that surface scope interpretations will be obtained if it is the first interpretation that is able to be verified.⁶ On this experiment, if it is the case that surface scope interpretations genuinely obtain an advantage, we predict that participants will stop on the first cup, at approximately the same rate at which adults persisted to verify the surface scope interpretation in previous experiments.

Alternatively, previous results may be explained by a desire for adults to wait until both interpretations have been revealed before making a judgment. In previous experiments, this would have resulted in participants refraining from judgment until the surface scope interpretation was verifiable.⁷ If it is the case that participants refrain from making a decision until all of the interpretations can be verified, then, in this version, we predict that adults will refrain from making judgments until they can

⁶Our gratitude goes to Kamil Deen for pointing out this prediction and suggested remedy.

⁷If this is the case, it is mysterious why the T/F judgments given by the participant align with the last interpretation that is verifiable.

verify the inverse scope interpretation. In this case, we could expect approximately 70% inverse scope interpretations.

Participants

20 undergraduate students at the University of Maryland, College Park, participated in this experiment. Two participants were replaced in the design for failure to obtain either interpretation on more than half of the target trials. Participants received either monetary compensation or course credit for their participation. All participants were English native speakers over the age of 18.

Results

Overall, adults obtained the inverse scope interpretation on 36.6% of trials. 10/20 participants obtained surface scope interpretations on all trials. The remaining 10 participants obtained the inverse scope interpretation on 68.3% of trials. The rates of inverse scope interpretations on this task do not differ from IVT: specified domain ($t(38)=1.26$, $p=.21$).

A total of 8 (out of 120) target items were excluded from analysis because they could not be interpreted under either interpretation. 5 of these were on the first trial. Participants obtained the correct response on 100% of unambiguous control trials.

Discussion

The rate of inverse scope interpretations obtained in this experiment does not differ significantly from the rate of inverse scope interpretations obtained in previous IVT experiments. Therefore, it appears that surface scope interpretations are obtained, regardless of which interpretation can be verified first. These results further confirm

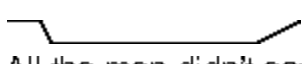
the finding that there exists an advantage for surface scope interpretations before disambiguating situational evidence is observed.

2.1.6 Experiment Five: IVT: Written

In all previous experiments, the target sentence has been verbally presented by the experimenter. It is possible then, that there are prosodic cues in the experimenter's presentation that bias the surface scope interpretation. Therefore, we must be assured that our results hold in the absence of any potentially disambiguating prosodic information.

Jackendoff (1972) claimed that for strings like (52), a falling intonation conveys a surface scope interpretation (53a), and a rising contour conveys an inverse scope interpretation (53b).

(52) All the men didn't eat

- a)  All the men didn't eat
- b)  All the men didn't eat

(53)

Presumably, these intonation contours, if used by the experimenter, may affect interpretation, under the assumption that adults can utilize prosodic information in ambiguity resolution.

Leddon et al. (2003) presented some evidence that adults may not produce or use intonation contour for disambiguating scopally ambiguous strings. They presented two experiments. In the first experiment, adults were instructed to read a storybook to a child that contained scopally ambiguous strings, such as (54). These strings were

embedded in a situation which clearly supported either the surface scope or inverse scope interpretation.

(54) Every bunny didn't jump over the fence

Adults did not produce an intonational contour that was predictive of interpretation, as compared to a pronoun control, where situation was predictive of contour. In a second experiment, Leddon et al. took samples of sentences produced from experiment one that contained a clearly disambiguating intonational contour, and presented them to adults. Adults interpreted these strings with inverse scope interpretations about 59% of the time, regardless of the intonational contour of the string.⁸ These results suggest that adults are poor at using rising and falling pitch to disambiguate scopally ambiguous strings.

Despite this finding, it is still possible that some aspect of the experimenter's verbal presentation is responsible for the high number of surface scope interpretations. It could be the case that the neutral prosody the experimenter used (as an attempt to not bias either interpretation) could contribute to surface scope interpretations. Or, there may be some other prosodic information (aside from the rise or fall on the last word), not tested by Leddon et al. that leads to interpretive biases. Finally, the data presented by Leddon et al. was child-directed speech, so it is possible the findings may not generalize to adult-directed speech.

It is important to replicate our previous results in a task where the stimuli are presented visually, to remove any prosodic cues. Of course, it is quite possible that when adults read a sentence, they assign the sentence 'silent prosody'. Fodor (2002) proposed that in reading experiments, the parser favors the syntactic analysis that

⁸Notice, these results suggest the bias toward interpreting scopally ambiguous strings with the inverse scope interpretation is not as strong as found by Musolino et al. (2000).

is associated with the most natural prosodic contour of the string. In this way, the silent prosody assigned to the string can influence ambiguity resolution. For the case of scope ambiguity resolution, we must consider what the most natural prosodic contour of the string is. There seem to be two possibilities. It may be the case that the natural contour is the one associated with the most common interpretation used in conversation, that is, the inverse scope interpretation. Alternatively, if the results by Leddon et al. are informative, it may be the case that there is no ‘natural’ prosody, as adults appear able to associate either interpretation to an intonational contour. In any case, there is no reason to believe that the silent prosody of scopally ambiguous strings is the one associated with surface scope interpretations. Therefore, if in this experiment, we continue to find an advantage for surface scope interpretations, we can eliminate all types of prosody as a source of the advantage.

Materials and Methodology

The cards, target sentences and unambiguous controls for this task remain the same as used for IVT: specified domain. However, the stimuli were presented in a written format. Participants were told that they would read a string, and would need to judge whether that sentence was true or false based on information on the card. The participants were instructed about the rules of the task the same way as in previous experiments. They were asked to read silently, but were not restricted from reading aloud if they desired. The remainder of the description of the task remains identical to previous experiments.

All trials are scored according to the previously mentioned criteria.

Target and unambiguous control items were mixed in a pseudo-random order. The order of presentation of items was counterbalanced. Two lists were created, with the

second list presenting target and control items in the reverse order of the first list. Half of the participants received list one, and the other half received list two. The entire session lasted approximately 10 minutes.

Predictions

If it is the case that prosodic information plays little role in the participant's interpretations, then we predict that the results of this experiment would mirror that of previous experiments. Specifically, we predict high rates of surface scope interpretations.

However, if it is the case that the experimenter's prosody biases participants toward surface scope interpretations, then we may expect higher rates of inverse scope interpretations in this experiment as compared IVT: specified domain.

Participants

10 undergraduate students at the University of Maryland, College Park, participated in this experiment. No participants were excluded. Participants received either monetary compensation or course credit for their participation. All participants were English native speakers over the age of 18.

Results

Across all target items, participants obtained the inverse scope interpretation 12.7% of the time. 2 trials were excluded from analysis for failing to indicate either interpretation. Performance on unambiguous control items was 100% correct.

These results do not differ from the results obtained in IVT: specified domain ($t(28)=-.65$, $p=.51$).

Discussion

In this experiment, participants obtained low rates of inverse scope interpretations. This shows that, even when the stimuli are presented in a visual format, an advantage for surface scope interpretations persists. Therefore, we can be confident that our previous results are not a result of a prosody that biased the surface scope interpretation.

We have presented a series of experiments that suggest that there exists an advantage for surface scope interpretations, even before the situation provides disambiguation cues. In the next experiment, we turn to investigating how the resolution of scope ambiguity differs from the resolution of attachment ambiguity with respect to the number of alternatives under consideration.

2.1.7 Experiment Six: IVT: Structural Ambiguity

We have claimed that this series of IVT experiments demonstrates that there exists an advantage for surface scope interpretations. This finding is part of an investigation to pinpoint the source of the advantage for surface scope interpretations that is found in adults, evident from this data, and also children, as evident in previous research. The larger focus of our investigation is to determine how LF parsing differs (if it does at all) from s-structure parsing. Therefore, it is crucial to investigate a case of s-structure ambiguity for comparison. Furthermore, in order to be confident that our findings are informative about the nature of scope ambiguity (and not simply all types of global ambiguity), we must investigate another type of global ambiguity.

One case of global s-structure ambiguity is attachment ambiguity, as in (55).

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



Figure 2.8: Example of Attachment Ambiguity

In the scenario in Figure (2.8), (55) 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 case of global s-structure ambiguity, analogous to the scope ambiguity case reviewed above.

There has been some work conducted with this construction in both adults (Tanenhaus et al., 1995) and children (Trueswell et al., 1999). Trueswell et al. (1999) investigated strings like (55), in either a one referent context (only one frog), or a two referent context, as is shown in Figure (2.9). Because Trueswell et al. were investigating the effects of the number of referents on eyegaze, they were investigating a scenario in which (55) is only temporarily ambiguous.

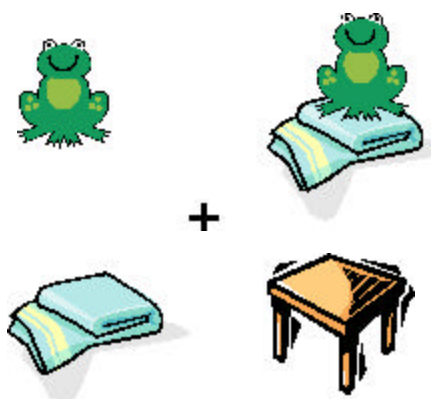


Figure 2.9: Example from Trueswell et al. (1999)

At the point of hearing *put the frog on the towel*, the string is ambiguous. The string 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 partic-

ipant should pick up *the frog that is on the towel*, and wait for an instruction about where to place it. As soon 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*.

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). However, children's eyegaze patterns suggested that they were unable to make use of situational information to disambiguate globally ambiguous strings.

We want to determine whether adults, with strings containing attachment ambiguity, select an interpretation before disambiguating information is encountered. To test this, we created an IVT that contained strings with PP attachment ambiguities, as in (56).

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

(56) is ambiguous between an analysis treating the PP *on the rug* as attached to NP (which will be called the 'NP-attached reading') or VP (which will be called the 'VP-attached reading'), as shown in (57).

- (57) 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 (57a) can be paraphrased as *pick up the cat and put it on the sofa that's on the rug*, whereas the VP attached PP (57b) 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 both alternatives are available until the situation disambiguates, then we expect adults to pick whichever interpretation for which there is sufficient information first.

Materials and Methodology

As with other IVTs, participants are presented with a series of cards, each card containing four scenes that are each hidden underneath a cup. Instead of making a truth-value judgment, participants are instructed to perform the indicated action as soon as possible, lifting the fewest number of cups. Some of the items in the scenes 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 Figure (2.10), both cats would be movable.

This experiment contained two conditions: the NP-attached and VP-attached conditions. Across both conditions, for a target sentence such as (58), the first picture is a cat, and the second cup reveals a rug.

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

In the NP-attached condition, the NP attached reading could be implemented on the third cup, as in Figure (2.10). (b) reveals a sofa on a rug, so an appropriate response is to place the lone cat on the sofa.

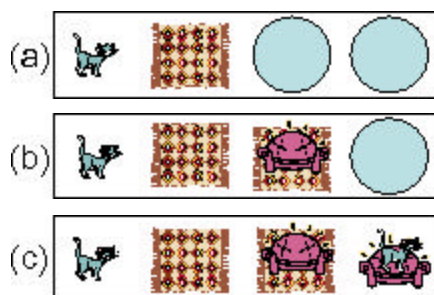


Figure 2.10: IVT: Syntactic Ambiguity, NP-attached Condition

The VP-attached condition makes the VP attached reading available on the third cup, as in Figure (2.11). (b) reveals a cat on a sofa, so an appropriate response is to take that cat and place it on the rug.

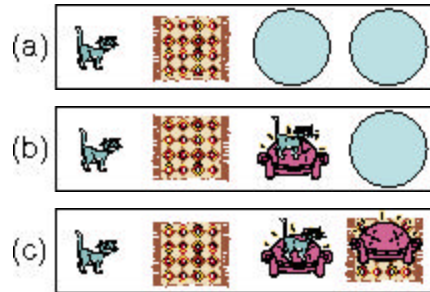


Figure 2.11: IVT: Syntactic Ambiguity, VP-attached Condition

Notice that in both conditions, the alternate interpretation can be obtained by lifting the final cup (Figure (2.10c) and Figure (2.11c)). This way, if a participant has only one interpretation available to him/her, the participant will be able to execute the action that corresponds to that interpretation.

Each participant received 6 target trials, 3 trials per condition. Each participant additionally received two warm-up trials, as an introduction to the task, and four unambiguous control trials. Unambiguous control trials were similar to the target trials, but utilized target sentences with relative clauses that force an NP (59) or VP (60) attachment.

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

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

The card for (59) is depicted in Figure (2.12).



Figure 2.12: IVT: Syntactic Ambiguity, NP-attached Filler

Trials were scored on both the action, and the cup the participant stopped on. If the action and cup number did not correspond to one of the interpretations, that trial was not scored.

Target and unambiguous control trials were mixed in a pseudo-random order, including the randomization of target conditions. Order of presentation was counter-balanced. Two lists were created, with the second list presenting target and control items in the reverse order of the first list. Half of the participants received list one, and the other half received list two. Each string was verbally presented by the experimenter with a prosody that did not favor either interpretation. The entire session lasted approximately 15 minutes.

Predictions

If the IVT, as a task, restricts participants from considering two interpretations, then we expect participants to obtain one interpretation across conditions. That is, condition would not correlate with the participant's interpretation.

If it is the case that the early disambiguation that we have seen in previous experiments is specific to scope ambiguity, then we may expect that with other types of global ambiguity, both interpretations remain available until a situational disambiguation point. Therefore, in an IVT where two interpretations are ultimately available, we predict that adults will stop at whichever interpretation that is available to be implemented first. 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 condition, we expect participants to pick up the lone cat.

Participants

20 undergraduate students at the University of Maryland, College Park, participated in this experiment. No participants were excluded from analysis. Participants received either monetary compensation or course credit for their participation. All

participants were English native speakers over the age of 18.

Results

We find that condition is highly predictive of interpretation. For each condition, we calculate the % NP attached readings. In the NP-attached condition, participants obtain 87% NP attached readings. In the VP-attached condition, adults obtain 21% NP attached readings, a significant difference between conditions ($t(19)=7.36$, $p < .001$). Therefore, the condition is highly predictive of the interpretation obtained by the participant. Only 4/20 participants obtained the same interpretation across conditions. Five trials were excluded from analysis due to failure to unambiguously correspond to either interpretation. These are divided as follows. On one trial, the participant took the cat that was already on the sofa, and moved it to the other sofa. On four trials, participants made ‘hopping errors’: taking the lone cat, moving it to the sofa, then moving it to the lone rug. 3 of these responses came from one participant. These types of errors have been noted in children as well (Trueswell et al., 1999).

On unambiguous controls, participants performed the proper movement, while stopping at the appropriate cup, 100% of the time.

Discussion

In this experiment, we found that the condition was highly predictive of interpretation. This suggests that both of the alternative s-structure representations of the target sentence are maintained until the disambiguation point, as participants are able to obtain either interpretation depending on the order of presentation of the pictures.

This result suggests that our findings concerning scopally ambiguous strings are unique to LF ambiguity. We can be confident that the IVT is capable of showing the dual maintenance of alternatives, and this dual maintenance of alternatives does not arise when participants encounter scopally ambiguous strings. Therefore, we conclude that there exists an advantage for surface scope interpretations in scopally ambiguous strings.

2.1.8 Experiment Seven: Speeded Forced Choice Task

Thus far, we have shown that within the IVT, adults obtain mainly surface scope interpretations. These results have led us to suggest that there exists an advantage for surface scope interpretations at some stage of processing. In conversation, adults prefer the inverse scope interpretation of strings with a subject universal quantifier and negation, so we infer that adults are able to make use of discourse information to revise their initial interpretation. However, in the IVT, there is a lack of rich discourse information, and therefore, no revision occurs. If it is the case that the IVT has prevented revision, then we should be able to find another task in which adults are unable to revise their initial interpretation. Assuming that the revision process requires time, one may expect that a time limit on interpreting the relevant strings may restrict an adult's ability to revise. If this is the case, we will have support that the surface scope advantage is more general than the IVT, and that surface scope interpretations are the initial interpretation obtained. Therefore, we turn to a speeded forced choice task.

Materials and Methodology

In this experiment, participants are presented with a story, and asked to complete a sentence by selecting an item from the story. An example completion prompt is

shown in (61).

(61) Every dwarf didn't paint the barn that's...

The preceding situation contains two barns: one that *no dwarves painted*, and one that *not all of the dwarves painted*. If the participant completes the target sentence with the barn that none of the dwarves painted, then this is evidence they obtained the surface scope interpretation. If the participant completes the target sentence with the barn that some of the dwarves painted, then we infer they obtained the inverse scope interpretation of the target sentence.

In this experiment, participants are presented with a picture on the computer screen, as in Figure (2.13), accompanied by an auditory explanation, below. The participant is told that their task is to complete a sentence about the story by choosing between one of two alternatives.

Here, there is a red, blue and green dwarf, with their cans of spraypaint. The farmer has pink spraypaint. There is a barn that the cow lives in, and a barn that the pig lives in. It looks like the red and blue dwarves spraypainted the cow's barn, but not the green dwarf. It doesn't look like any of the dwarves spraypainted the pig's barn, so the farmer finished the job.

In this case, *the pig's barn* corresponds to the surface scope interpretation, because no dwarves spraypainted it. *The cow's barn* corresponds to the inverse scope interpretation, because not all dwarves spraypainted it. Therefore, if the participant selects the pig's barn as a completion of (61), we infer they obtained the surface scope interpretation of the string. If the participant selects the cow's barn, we infer that they obtained the inverse scope interpretation.



Figure 2.13: Speeded Forced Choice task, story slide

There are three types of control sentences. Control sentences either contained a universal quantifier and no negation, as in (62), a definite article and no negation, as in (63), or a definite article and negation, as in (64).

(62) Every kangaroo wallpapered the wall that's ...

(63) The blue kangaroo wallpapered the wall that's ...

(64) The blue kangaroo didn't wallpaper the wall that's ...

All types of control sentences are accompanied by a picture (an example is shown in Figure (2.14)), and an auditory explanation, exemplified by the story presented below.

Here, there is a brown, yellow and blue kangaroo, with their pieces of wallpaper. The snail has grey wallpaper. There is a red wall and a purple wall. It looks like the brown and yellow kangaroos wallpapered the red wall, but not the blue kangaroo. The snail helped, too. It looks like the brown, yellow and blue kangaroos all wallpapered the purple wall.

The presentation format remained the same as for targets. This experiment contains two conditions, a speeded and a non-speeded condition. In the speeded condition, participants are told that their task is to complete a sentence about the story by selecting from one of two pictures (by pressing either the F or J key on the keyboard).

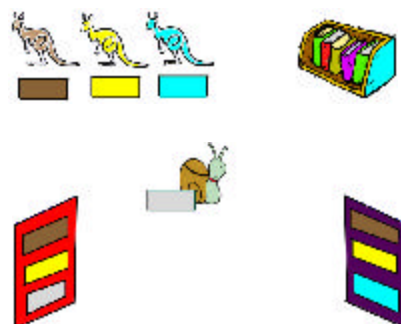


Figure 2.14: Speeded Forced Choice task, control slide

At the end of the story, participants are shown a screen, as in Figure (2.15), which contains both barns, labeled according to their distinctive features.

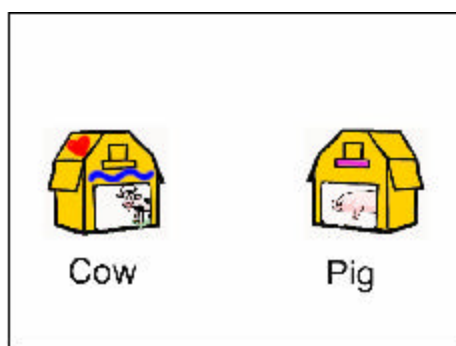


Figure 2.15: Speeded Forced Choice task, choice slide

This slide is presented for 1400ms, and then is replaced by a fixation point (for 750ms). Then, the final slide reappears, and the target sentence is played, either as in (65) or (66). The target sentences were divided evenly between these two types of prompts.

(65) Every dwarf didn't spraypaint the *

(66) Every dwarf didn't spraypaint the barn that belongs to the *

The * represents a tone. Participants are told they should select the proper picture as soon as possible after the tone, by pressing either the key designated for the left picture or the key designated for the right picture. The timeline is represented in

Figure (2.16).

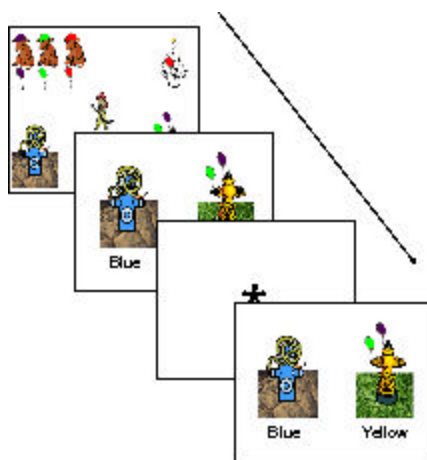


Figure 2.16: Speeded Forced Choice task, presentation timeline

This condition contained 10 target and 14 control items: 8 with the universal quantifier and no negation, 3 with a definite article and no negation and 3 with a definite article and negation. Items were presented in a random order using Psyscope. Because the response is a button press, no items were excluded from analysis in this condition.

The non-speeded condition utilized 6 target stimuli from the speeded condition.⁹ Participants were not presented with the final choice slide, and were presented with the string in (66), without a beep. Participants were given as much time as needed to answer, and responded verbally to the experimenter, without a button press response.

This condition contained 6 target items and 3 control items (all with the universal quantifier and no negation). Each participant received one of the two possible presentation lists.

⁹This condition contains a different number of trials to align with a different experimental condition, which is presented in Section 4.2.

In the non-speeded condition, responses were scored according to which ‘barn’ the participant referred to. For example, the following types of responses would all be scored as referring to the cow’s barn: pointing to the cow’s barn, answering ‘*the cow*’ or providing a response which unambiguously referred to the cow’s barn, such as ‘*the barn that two of the dwarves painted*’. A response that did not unambiguously refer to either barn, as in the case of a naming error that could not be resolved, such as ‘*the horse’s barn*’, were excluded from analysis.

Across conditions, the pictures were counterbalanced for which side of the screen the item consistent with the surface scope interpretation appeared on. The auditory explanation always described the story in a counterclockwise fashion, beginning with the characters in the upper left hand corner.

Predictions

If it is the case that adults initially give preferential consideration to the surface scope interpretation, and require the integration of discourse and situational information to revise to the inverse scope interpretation, then we predict that adults will obtain higher rates of surface scope interpretations in the speeded condition over the non-speeded condition. This would reflect a lack of time available to integrate discourse information in a way consistent with the inverse scope interpretation.

Alternatively, it may be the case that the time constraints in the speeded condition are insufficient in preventing adults from integrating the discourse information required to obtain inverse scope interpretations. If this is the case, then we predict very little difference between conditions.

Participants

40 undergraduate students from the University of Maryland, College Park, participated in this experiment, with one participant replaced in the design due to less than 70% correct performance on control trials. Participants were divided into either the speeded or non-speeded condition. Participants received either monetary compensation or course credit for their participation. All participants were English native speakers over the age of 18.

Results

In the speeded condition, adults obtained the inverse scope interpretation on 18.5% of trials. No trials were excluded from analysis. Performance on control items was 90% correct.

In the non-speeded condition, adults obtained the inverse scope interpretation on 40% of trials. If a participant gave both answers, this was scored as an inverse scope interpretation, because we are investigating the adult's ability to obtain the inverse scope interpretation. This occurred on two trials total. 10/20 participants obtained the surface scope interpretation on all trials. The remaining 10 participants obtained the inverse scope interpretation on 86.6% of trials. No trials were excluded from analysis. Performance on control trials was accurate 98.3% of the time.

The two conditions are significantly different ($t(38)=2.02$, $p < .01$).

Discussion

We find that in a task that imposes time pressure on resolving ambiguity, adults obtain largely surface scope interpretations. The results from the IVT suggested that adults

have an advantage for surface scope interpretations at some stage of processing. However, because adults adhere to inverse scope interpretations in conversation (Musolino and Lidz, 2006), we infer that adults must be able to revise their interpretation. For the results from the IVT, we claimed that the lack of rich discourse information prevented adults from revising. If this claim is to be more general, that is, if revision can be prevented and adults will adhere to surface scope interpretations, then we should hope to replicate this result in other tasks. With the speeded sentence completion task, we showed that when adults must resolve a scopally ambiguous string under time pressure, then adults adhere to the surface scope interpretation. This is in contrast to the non-speeded condition, presumably where adults had available the time needed to revise their interpretation. This finding further supports the advantage for surface scope interpretations in adults.

2.2 Summary

We have shown a series of experiments that suggest that, different from s-structure ambiguity, adults begin to resolve scope ambiguity even before encountering disambiguating information. A summary of the experiments in this section (that directly test strings containing the subject universal quantifier and negation) is shown in Table 2.1.

Table 2.1: Summary of IVT Experiments

Experiment	UQ	Specified	Reverse	Written	Speeded Choice
% Inverse Scope	22.9	21.1	36.6	12.7	18.5

On the path from hearing a string to verification, it appears that adults only maintain one alternative, suggesting there exists a step in the sentence comprehension process that resolves ambiguity before situational disambiguation.

Under the parsing hypothesis, these results are easily captured. The surface scope interpretation is the initial interpretation obtained by the parser. The inverse scope interpretation requires revision, and requires the presence of discourse information to trigger this revision. In the IVT, where discourse information is absent, there is no revision to the inverse scope interpretation, and therefore, the parser resolves the ambiguity according to its initial interpretation. The time pressure of the speeded sentence completion task is another experimental way in which revision from the surface scope interpretation can be prevented. That is, when time is restricted, there is no revision from the parser's initial interpretation. This hypothesis can also explain why s-structure ambiguities are handled differently from LF ambiguities. For example, The PF/LF similarity version of the hypothesis posits that parsers interpret strings rapidly, and because many strings are unambiguous, the parser makes the assumption that subject quantifiers can be interpreted in their case positions. Therefore, the parser is confident committing to one interpretation, resulting in scope ambiguity being handled serially. In the case of PP attachment ambiguity, attachment preferences are determined by lexical items, and therefore the parser cannot make such strong predictions about the logical form based on the s-structure alone. Thus, s-structure ambiguities may not have a strong parsing preference associated with the string, allowing flexibility in interpretation. One piece of data remains unexplained on this account. In the non-speeded condition of the forced choice task, adults obtained inverse scope interpretations only 50% of the time. If it is the case that surface scope interpretations are a parsing default, this suggests that the discourse in the sentence completion experiment forces revision 50% of the time. It is an open question, then, why, if inverse scope interpretations are 'preferred', the rate of revision in this case is so low.

These results can also be explained under the extra-linguistic hypothesis. Recall that the extra-linguistic hypothesis posits that the parser makes both interpretations equally available, and that factors external to the parser are responsible for the advantage for surface scope interpretations. This hypothesis naturally explains the results that appear mysterious under the parsing hypothesis. That is, in a forced choice task, where neither interpretation is more supported in the discourse than the other, adults are at chance with obtaining the inverse scope interpretation. This is perfectly described by a system where both interpretations are equally available, and in the absence of discourse biases, both interpretations remain equally available for selection. The remainder of our results follow less naturally. Here, we will put forth an outline of an extra-linguistic hypothesis, which will be elaborated on in further detail in subsequent chapters. Let us assume that there exists an extra-linguistic factor that favors the surface scope interpretation (we will later argue that this factor is the ease of verification procedure available). In the case where the parser makes multiple interpretations equally available, this factor will work to select the surface scope interpretation. In the case of the sentence completion experiment, the participant is involved in a rich discourse situation, so the bias toward inverse scope interpretations in conversation mitigates this effect. However, in the IVT, there is no conversational discourse, leaving the extra-linguistic pressures to surface scope interpretations to arise. Importantly, according to this hypothesis, the advantage for the surface scope interpretations can be derived, and not attributed to a parser bias.

These two hypotheses view the nature of the parser differently. If it is the case that there exist parsing preferences specific to scope ambiguity resolution, which do not arise in the case of s-structure ambiguity resolution, then this suggests that the parser treats these two types of ambiguity differently. That is, this could be taken as evidence that the parser makes reference to s-structures and logical forms as different

levels of representation. However, if the surface scope advantage can be described by the extra-linguistic hypothesis (that is, there are forces external to the parser that are responsible for the selection of surface scope interpretations), then this hypothesis leaves open the possibility that the parser handles s-structures and logical forms in the same way.

Recall that we began investigating the adult system as a method of determining how the child and adult systems are alike. We have found a task that restricts adults to the surface scope interpretation. It seems to be the case that children, generally, appear restricted to the surface scope interpretation. Under the view that these two observations are connected, we are interested in whether we have found a task that allows us to mimic, in adults, the process by which children resolve scopally ambiguous strings. In the next section, we investigate priming in adults, as a means of investigating whether the same factors underlie the surface scope interpretation in adults and children.

Chapter 3

Priming

We have shown results from two tasks, the IVT and the speeded forced choice task, in which adults obtain high rates of surface scope interpretations. Previous research shows that children, at least in experimental situations, obtain high rates of surface scope interpretations. We first investigate the degree to which the surface scope interpretations obtained by adults mimic the pattern we observe in children. If it is the case that we have ‘turned adults into children’ in our tasks, then we expect adults obtain inverse scope interpretations in the same experimental manipulations that allow children to obtain inverse scope interpretations. We investigate two experimental manipulations that increase access to inverse scope interpretations in children: the presence of a positive lead in sentence (Musolino and Lidz, 2006) and priming (Viau et al., 2006). We show that adults, in the IVT, obtain higher rates of inverse scope interpretations in these same experimental manipulations. This suggests that the same factor restricts adults interpretations in these tasks and children’s interpretations generally.

Second, we investigate the nature of priming. Previous research on priming of scope interpretations has focused mainly on children, and centered on the investigation of

the priming of inverse scope interpretations. We add to the investigation of priming in two ways: one, by investigating priming effects in adults, and two, by investigating the priming of surface scope interpretations. We show that surface scope interpretations can be primed in adults. This suggests that inverse scope interpretations have no privileged status with respect to priming. In a second priming experiment, we find that priming is asymmetrical, with only dispreferred representations being primed.

3.1 Turning Adults into Children

We have found that in certain tasks, adults behave like children, in that they appear to access mostly surface scope interpretations. Ultimately, we are interested in whether children's restrictions to surface scope interpretations reflect the initial stages of adult-like sentence processing. That is, to what degree do children follow an adult-like procedure for sentence processing, but halt at an earlier point in the process? One way of investigating the similarity between children and adults in the IVT is to see whether adults in the IVT pattern in other ways like children. If the same experimental manipulations that assist children also allow adults to gain access to inverse scope interpretations in the IVT, this would be evidence that the IVT imposes in adults limitations similar to those found generally in children.

3.1.1 Experiment Eight: IVT: Preceding Sentence

Musolino and Lidz (2006) found that children obtained high rates of inverse scope interpretations when the presence of negation was made felicitous in the target sentence, as in (67).

(67) Every horse jumped over the log, but every horse didn't jump over the fence

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

In a scenario where two out of three horses jump over the fence, children obtained the inverse scope interpretation of (67) 60% of the time, significantly higher than the children in the control condition, who received target sentences like (68) and obtained the inverse scope interpretation only 15% of the time. These results are shown in Figure (3.1).

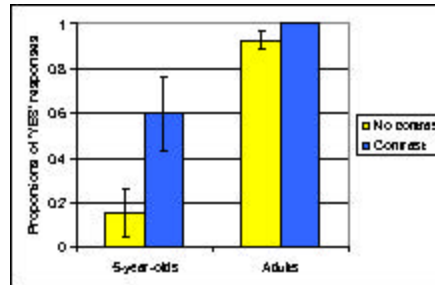


Figure 3.1: Positive lead-in graph from Musolino and Lidz (2006)

These results show that the positive lead-in sentence improved children's access to the inverse scope interpretations.

If it is the case that the surface scope interpretations obtained by adults in the IVT result from the same factors present in children more generally, then we may expect this same manipulation to improve adults' access to inverse scope interpretations. This finding would not reveal the mechanisms underlying such interpretations, simply that children and adults are parallel in this regard. In this experiment, we perform an IVT where adults receive the same type of positive lead-in as in Musolino and Lidz.

Materials and Methodology

The procedure for presentation of the IVT to the participant remains the same as all other IVT experiments, with one minor change in directions. Participants are told that the experimenter will utter two sentences about each card, but that their task is to judge the truth/falsity of the second sentence only. The participant is told that

the first sentence will always be true, and is just a lead-in to the statement of interest.

In this experiment, we used target sentences like (69).

(69) Every cow has spots, but every cow doesn't have a hat

We used the same cards as for IVT: specified domain. Therefore, each participant received 6 IVT target trials, plus 5 unambiguous control trials. A positive lead-in string was added to the unambiguous control items, as shown in (70), so that the target and unambiguous control items remained maximally similar.

(70) Every sheep is fluffy, and every sheep has a sweater.

Both target and unambiguous control items were scored according to the previous criteria.

The order of presentation of items was counterbalanced. Two lists were created, with the second list presenting target and control items in the reverse order of the first list. Half of the participants received list one, and the other half received list two. Target sentences were presented in a neutral prosody that did not bias either interpretation. The entire session lasted approximately 10 minutes.

Predictions

We know that adults obtain a high rate of surface scope interpretations in the IVT, which we have suggested reflects adults' inability to integrate the discourse information required to obtain an inverse scope interpretation. If it is the case that the task mirrors in adults the same deficits that are generally present in children, then we expect the existence of a positive preceding string to increase rates of inverse scope interpretations obtained by adults in this task.

Alternatively, if the high rate of surface scope interpretations we observed in the IVT results from a different source than in children, then we may expect the preceding sentence condition to have no effect in eliminating the difficulty associated with obtaining inverse scope interpretations in adults.

Participants

10 undergraduate students at the University of Maryland, College Park, participated in this experiment. No participants were excluded from analysis. Participants received either monetary compensation or course credit for their participation. All participants were English native speakers over the age of 18.

Results

In this experiment, adults obtained the inverse scope interpretation in 56% of target trials. This is significantly higher than the percentage of inverse scope interpretations obtained in IVT: specified domain ($t(28)=2.40$, $p < .03$). A total of 2 trials were excluded from analysis due to failure to correspond to either interpretation.

Participants responded correctly on 100% of unambiguous control trials.

Discussion

These results suggest that the contrast provided by a positive lead-in sentence also works in the IVT to allow adults to recover from the surface scope interpretation. This is one piece of evidence that the factors that restrict children to surface scope interpretations also restrict adults to surface scope interpretations in certain discourse environments. This suggests that if we can identify the factors that encourage adults to arrive at inverse scope interpretations, then we can also identify the limitation

present in children. If it is the case that children and adults arrive at surface scope interpretations for the same reasons, then we also predict that priming, which has been shown to help children obtain inverse scope interpretations in children, would also assist adults. Let us now turn to the investigation of priming in the IVT.

3.1.2 Experiment Nine: IVT: Prime

We found that a positive lead-in sentence, which helps children obtain inverse scope interpretations, also aids adults in obtaining inverse scope interpretations in the IVT. Another experimental manipulation that helps children obtain inverse scope interpretations is priming. The same logic here, holds. If it is the case that adults and children are both aided by priming, then this is strong evidence we are tapping into the same mechanism.

Viau et al. (2006) found that strings of the form in (71) prime the inverse scope interpretation of strings like (72) in children. Without priming, children obtain mostly surface scope interpretations of sentences like (72).

(71) Not every bug hid behind the tree

(72) Every bug didn't hide behind the tree

Viau et al. tested two groups of children. The first group was presented with 6 trials of sentences like (72), with appropriate fillers. In this condition, children obtained the inverse scope interpretation 22.25% of the time on the first three trials, and 38.8% of the time on the last three trials. The second group of children had 3 sentences of the form (71), followed by 3 sentences of the form (72). In this condition, children obtained *not every* interpretations 83.3% of the time in the first three trials. In the last three trials, children obtained the inverse scope interpretation 80.58% of the time. The asymmetry between inverse scope interpretations obtained on the last

three trials between conditions demonstrates a strong priming effect. These results are shown in Figure (3.2).

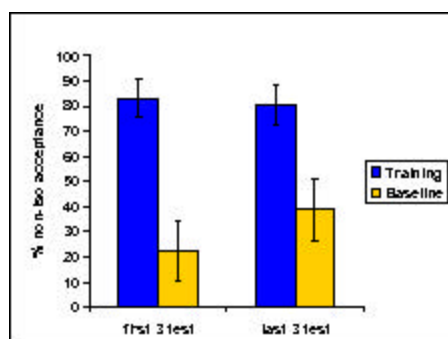


Figure 3.2: Prime graph from Viau et al. (2006)

It appears that a string which unambiguously leads children to accept the *not every* interpretation also helps them access this same scope relation even when the string is no longer unambiguous.

Given that adults obtain a high proportion of surface scope interpretations in the IVT, we want to determine whether priming assists adults in obtaining inverse scope interpretations. If so, this is another piece of evidence that the pressures underlying the advantage for surface scope interpretations are the same in children (generally) and adults (in the IVT and speeded forced choice task). This finding, would allow us to make further investigations into the common factors in children and adults that lead to the advantage for surface scope interpretations.

Materials and Methodology

We designed an IVT to test the effects of priming on interpretation. The experimental procedure is identical to that used in previous IVT experiments.

This experiment contained 9 target items. The cards were designed in the style of IVT: universal quantifier, where the quantified subject is hidden underneath the

cup.¹ The first and last three target items are of the form *every x doesn't...*, as in previous IVT experiments, and as in (72). The middle three items were of the form *not every x ...*, as shown in (71). This experimental design allowed us to obtain both a baseline measure of the adult's performance in the IVT (the first three items) and a measure of how this performance is effected after the prime (the last three items). This experiment contained the same style of unambiguous controls as in IVT: specified domain, with 4 *every* sentences and 4 *some* sentences.

All target and unambiguous controls were scored under the same criteria as previous experiments. Furthermore, for the *not every* trials, the expected response was to stop on the first cup, as it would be true that not all of the items had the specified property. However, some participants wanted to verify the implicature *not none*. These responses are coded.

The order of presentation of items was counterbalanced. Two lists were created, with the second list presenting target and control items in the reverse order of the first list. Half of the participants received list one, and the other half received list two. 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.

Participants

20 undergraduate students at the University of Maryland, College Park, participated in this experiment. No participants were excluded from analysis. Participants received either monetary compensation or course credit for their participation. All participants were English native speakers over the age of 18.

¹The experiment was designed this way for chronological reasons, not for methodological reasons.

Results

First, we can look at the first three target items, which are scopally ambiguous strings before the prime. Here, adults obtain the inverse scope interpretation 21% of the time. These results are consistent with previous IVT results.

The last three target items are scopally ambiguous strings that occur after the prime. Here, adults obtain the inverse scope interpretation 43.3% of the time. Performance on these three trials is statistically different from the first three trials ($t(19) = -2.73$, $p = .013$). These results are shown in Figure (3.3).

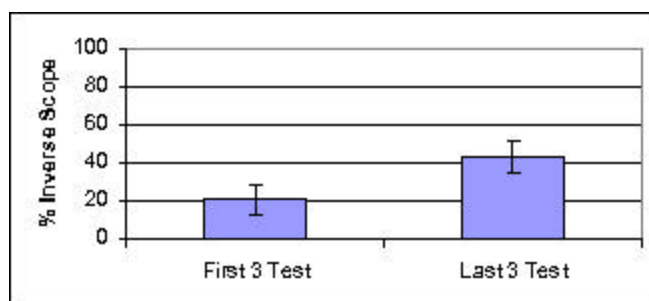


Figure 3.3: IVT: Prime, percent inverse scope by trial

A total of three items were excluded from analysis due to non-correspondence with any interpretation.

Of 57 total *not every* trials, participants stopped on the first cup and said ‘true’ 82% of the time (47 trials). There were 10 trials in which participants gave non-ideal responses. These are divided as follows. 1 trial was excluded for failing to indicate any possible interpretation (subject error). On two trials, the participant persisted to the last cup. On 7 trials, the participant turned over a second cup, and stopped. All except one participant had at least one ideal response to a *not every* trial. These seven trials are within the one-cup leniency afforded to target trials.

Across all unambiguous control trials, participants stopped on the appropriate cup on 100% of the time.

Discussion

We can conclude that *not every* sentences prime the inverse scope interpretation of scopally ambiguous strings containing the universal quantifier and negation. This finding is consistent with the child results. This serves as another piece of evidence that the limitations caused by the IVT mirror children's overall limitations.

Additionally, the *not every* trials allowed us to verify the validity of an assumption implicit in the IVT. We have assumed that if an adult obtains the inverse scope interpretation, they will stop on the first cup. If it is the case that adults obtain the inverse scope interpretation, but persist beyond the first cup, then our conclusions regarding the advantage for surface scope interpretations would be undermined. The *not every* trials require the same response as obtaining inverse scope interpretations, and participants stop on the first cup on 82% of trials, even on the strictest scoring criterion. When we score according to the one-cup leniency criteria used for all target items, then participants provide the expected response on 94% of trials. This confirms our assumption that obtaining the inverse scope interpretation results in the behavior of stopping on the first cup.

3.1.3 Intermediate Conclusions on the IVT

We have reviewed two experiments in which children obtain inverse scope interpretations with sentences with a universally quantified subject and negation. We found that these same experimental manipulations increased access to inverse scope interpretations for adults in the IVT. This suggests that the same mechanisms underlie

adults' and children's limitations in obtaining inverse scope.

These results do not raise any novel questions, but there is one issue from the child literature that should be taken up here. It is an open question why *not every* sentences prime the inverse scope interpretation of *every x didn't* sentences. It is true that both the *not every* sentences and the inverse scope interpretation of a scopally ambiguous string have the same logical form. But this is not a trivial description of priming. That is, one formulation of priming is that priming acts on a PF $\rightarrow$ LF process, requiring identical forms of PFs. However, these results suggest that this is not the case, as the *not every* sentences are of a different phonological form from the scopally ambiguous strings. Therefore, it must be the case that priming occurs over the logical form representation. It seems that in this case, priming occurs over representations, as the *not > every* scope relation is what is primed. Any account of priming that appeals to procedures (such as priming the operation of Quantifier Lowering) will fail to account for this data, because the *not every* sentences will not undergo this operation. In the next section, we turn to an investigation of the nature of priming.

3.2 On the Nature of Priming

Given a sentence with two interpretations, there are two prerequisites for the appearance of priming effects. First, one needs a prime sentence that effectively yields a given interpretation. In the previous experiment, *not every* sentences, which reliably obtained *not every* interpretations, served as prime sentences. Second, the interpretation that one is attempting to prime cannot be at ceiling in the task one is testing. That is, one could not prime inverse scope interpretations in adults in a TVJT because adults already obtain 99% inverse scope interpretations. In this case, we cannot

make a deep claim about the nature of priming, but simply that ceiling effects obscure the availability of priming.

Previous research has shown that inverse scope interpretations can be primed in English-speaking children. The priming of other interpretations and with other age groups has not been tested due to the unavailability of one or both of the prerequisites mentioned above. We have been unable to test the priming of surface scope interpretations in children because rates of acceptance of surface scope interpretations are already at ceiling. In adults, we have been unable to test the priming of inverse scope interpretations for the same reason. Finally, we have been unable to test the priming of surface scope interpretations in adults because we have lacked a suitable prime that guarantees access to the surface scope interpretation.

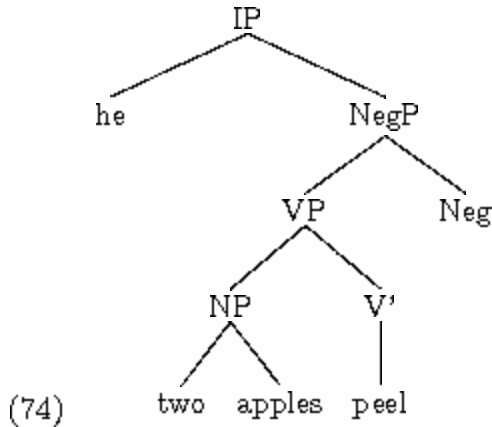
Given our previous inability to test these priming possibilities, let us outline three possible conclusions we might draw from the child data. One, we might conclude that any interpretation can be primed, and our failure to observe priming across the full range of interpretations is due to the absence of the proper constructions to test these effects. Two, we may conclude that only non-initial interpretations (such as inverse scope) can be primed. For this hypothesis to be the case, we must assume that the surface scope interpretation is the first interpretation obtained, and that obtaining the inverse scope interpretation requires revision. Third, we may conclude that LF priming behaves like lexical and s-structure priming, where it has been suggested that dispreferred interpretations display stronger priming effects (Scarborough, 1977; Ferreira, 2003). In the case of scope ambiguity, we may predict that because children prefer surface scope interpretations, only inverse scope interpretations can be primed. Adults, who prefer inverse scope interpretations can be primed only by surface scope interpretations.

With respect to the third hypothesis, it cannot be the case that priming occurs only when the prime is dispreferred with respect to the participant's general preference. We have shown in IVT: prime that adults, in the IVT, can be primed with the inverse scope interpretation. This argues against the idea that a person's general preference is the only information taken into consideration for calculating the dispreferred representation. Therefore, let us revise this hypothesis to state that the prime must be dispreferred with respect to the preceding environment. In the case of IVT: prime, the prime occurred after IVT trials, in which surface scope interpretations are preferred. Therefore, inverse scope priming can obtain. In the absence of an experimental preceding environment, the participant's general preference is used as the preceding environment. Because this is a delicate notion, let us draw an analogy to explain it. Priming effects, under this hypothesis, basically occur when something noticeably different happens with regards to interpretation. We can intuit this claim by appealing to our judgment of temperature conditions. For example, a person from Miami is likely to judge 50 degrees as chilly. However, we know that one's general experience with temperature (their 'preceding environment') does not solely determine our weather judgments. If this same person from Miami spends a week in the South Pole, then they are less likely to judge a subsequent 50 degree temperature as cold. That is, the short-term experience becomes used for calculating the 'preceding environment' by which surprise is measured. We claim that the same is the case, here.

Previous results do little to inform us about the nature of priming, leaving us with a number of open hypotheses. First, we will review an experiment that circumvents the problem concerning the priming of surface scope interpretations in children. Next, we will move to our own investigation of priming effects in adults.

We have seen how, in English, we are unable to prime surface scope interpretations in children due to ceiling effects. Lidz and Conroy (2007) tested priming in children who speak Kannada, a language which provides constructions that eliminate these obstacles. We presented children with strings that contained negation and a number in object position, as shown in (73), in a TVJT. This construction allows the same ambiguity as in English, either *it is not the case that he peeled two apples* or *there are two apples that he failed to peel*. Also similar to English, children prefer surface scope interpretations of strings like (73) in this task (Lidz and Musolino, 2002). The s-structure for the string in (73) is shown in (74).

- (73) avanu eraDu seebu orey-al-illa
 he two apple peel-inf-neg
 ‘He didn’t peel two apples’



- (75) avanu ii seebu-gaL-alli eradu orey-al-illa
 he prox apple-pl-loc two peel-inf-neg
 ‘He didn’t peel two from these apples’

With sentences like (73) (which we will call the zero-morpheme sentence, as the direct object contains no case morphemes), children obtained inverse scope interpretations 22.5% of the time. With the partitive construction (where the direct object is marked with a partitive morpheme), as shown in (75), children obtain the inverse scope interpretation 91.7% of the time. These constructions allow the perfect opportunity for the testing of priming, because we have both requisite types of prime

sentences, and have eliminated ceiling effects. If it is the case that priming can act on any interpretation, then we would predict that getting the partitive sentence first would increase inverse scope interpretations in the zero-morpheme case, and getting the zero-morpheme sentence first would increase surface scope interpretations for the partitive.

We tested 40 children, split into two conditions: Partitive-first and Zero-first. In the Zero-first condition, children received 3 zero-morpheme target sentences followed by 3 partitive sentences. In the Partitive-first condition, children received 3 partitive target sentences followed by 3 zero-morpheme sentences. The children in the Partitive-first condition showed an effect of priming, obtaining a higher rate of inverse scope interpretations in the zero-morpheme case than controls (93.2% vs. 22.5%). However, children in the Zero-first condition showed no evidence of surface scope priming when they received the partitive sentences. The 16.5% rate of inverse scope interpretations is similar to the control condition. These results are shown in Figure (3.4).

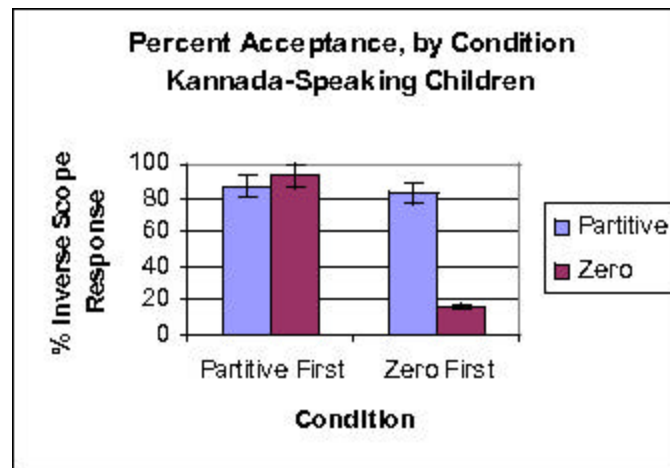


Figure 3.4: Kannada-prime graph, from Lidz and Conroy (2007)

These results suggest that priming, in children, is asymmetric. This finding bears on our current hypotheses. Recall, our first hypothesis was that priming occurred for any construction, regardless of preference, but that this observation was obscured

by ceiling effects. The Kannada case eliminated the ceiling effects, and found that only the inverse scope interpretation could be primed. These results are consistent with either of the two remaining priming hypothesis. It may be the case that only non-initial interpretations can be primed or that only dispreferred interpretations (the inverse scope interpretation for children, compared to their general surface scope preference) can be primed. Let us now turn to adults.

3.2.1 Experiment Ten: IVT: TVJT

In this experiment, we investigate the existence of a priming relation between the interpretations in the IVT, a task which induces surface scope interpretations in adults, and the interpretations in the TVJT, a task in which adults obtain only inverse scope interpretations. The IVT provides us with a surface scope prime sentence for adults, so we can test whether there exists an asymmetry in the priming of interpretations in adults. We find that only surface scope interpretations show an effect of priming, suggesting that only dispreferred interpretations relative to the preceding environment can be primed.

Materials and Methodology

In this experiment, we used a within-subjects design, with an IVT condition and a TVJT condition. Adults were introduced to the tasks as separate tasks. The IVT was introduced to the participant as in previous IVT experiments. In the TVJT condition, participants were told that they would watch a story, and at the end of the story, a puppet would attempt to describe what happened, but that the puppet may be right or wrong. Participants were told that their task was to tell the puppet whether it was right or wrong about its description of the story.

The IVT condition contained the 6 target items and 5 unambiguous control items from IVT: specified domain. The TVJT condition contained 3 target items, as exemplified in (76) (the final scene is pictured in Figure (3.5)), and one filler item.

(76) Every girl didn't catch a turtle

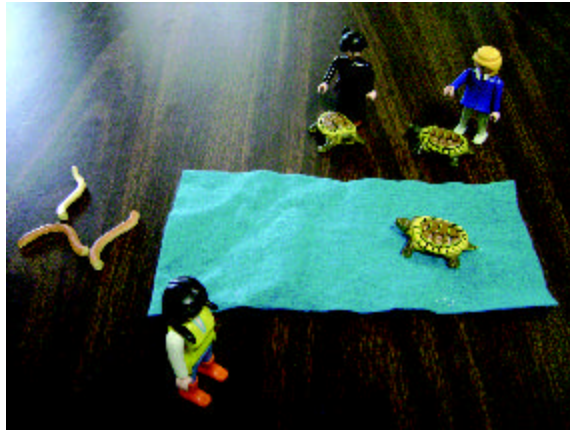


Figure 3.5: IVT: TVJT, TVJT example story

The story preceding the target sentence in (76) followed the design in Musolino et al. (2000), and is shown below.

In this story, there are three girls, who each consider catching snakes. They decide the snakes are too scary, and decide to catch some turtles from the pond. Each of the first two girls succeed in catching a turtle, but the third girl realizes that she can't swim, and does not catch a turtle.

In this story, the surface scope interpretation of (76), *none of the girls caught a turtle*, is made possible when the three girls consider catching snakes. This interpretation is made false when the two girls catch turtles. When the third girl fails to catch a turtle, the inverse scope interpretation of (76), *not all of the girls caught a turtle*, is made true.

The filler sentence followed a story of the same format, and could be altered dynamically by the experimenter to be either true or false, so that in the task, adults received a variety of true and false sentences. An example of a true filler sentence is shown in (77), and an example of a false filler sentence is shown in (78) for the story shown above.

(77) Some girls caught a turtle

(78) All girls caught a turtle

Target and filler items were presented in a pseudo-random order. Each sentence was verbally presented by the experimenter with a prosody that did not favor any specific interpretation.

Half of the participants received the IVT condition first, and the other half received the TVJT condition first. The entire session lasted approximately 15 minutes.

Predictions

The first hypothesis is that only non-initial interpretations can be primed. Because inverse scope interpretations require revision, this hypothesis predicts that only inverse scope interpretations can be primed. This account is most easily viewed as priming the revision procedure. That is, if it is the case that a participant gains experience revising their initial interpretation, then that revision is more likely to occur in subsequent analyses.² In this case, we expect inverse scope priming in conditions

²This hypothesis relies on the parsing hypothesis for the advantage for surface scope interpretations, which posits that surface scope interpretations are initial interpretations. If it is not the case that the inverse scope interpretation requires revision to obtain, then the prerequisite for this hypothesis would not be met. That is, if it is the case that the parser makes both interpretations equally available, and obtaining the inverse scope interpretation is the result of using discourse information to select that interpretation, then there is no way in which the process of revision from

when the IVT follows the TVJT, but we do not expect surface scope priming when the TVJT follows the IVT.

The second hypothesis is that only dispreferred interpretations can be primed, in an analog to what is found for lexical and s-structure priming. Recall that we must claim that the notion of ‘dispreferred’ must be calculated with respect to the preceding environment. Because the prime occurs at the beginning of the experiment, we posit that adults’ general preference, for inverse scope interpretations, will determine the preceding environment. If it is the case that priming occurs only when the primed interpretation differs from the interpretation in the preceding environment, then we predict priming for surface scope interpretations, but not for inverse scope interpretations. More specifically, we predict no inverse scope priming when the IVT follows the TVJT, but surface scope priming when the TVJT follows the IVT.

Participants

21 undergraduate students at the University of Maryland, College Park, participated in this experiment. No participants were excluded from analysis. Participants received either monetary compensation or course credit for their participation. All participants were English native speakers over the age of 18.

Results

First, we can look at the target items presented in the first condition to the participants. The participants who received the IVT first obtained inverse scope interpretations 33.3% of the time. This is consistent with previous experiments in the IVT. The participants who received the TVJT first obtained the inverse scope interpretation 96.6% of the time. This is consistent with previous results reported in the literature.

the surface scope interpretation could be primed.

Now let us turn to the investigation of the effects of priming. The participants who were given the TVJT first obtained the inverse scope interpretation in 8.4% of the IVT trials. The participants who were given the IVT first obtained the inverse scope interpretation 72.7% of the time in the TVJT trials. These results are shown in Figure (3.6).

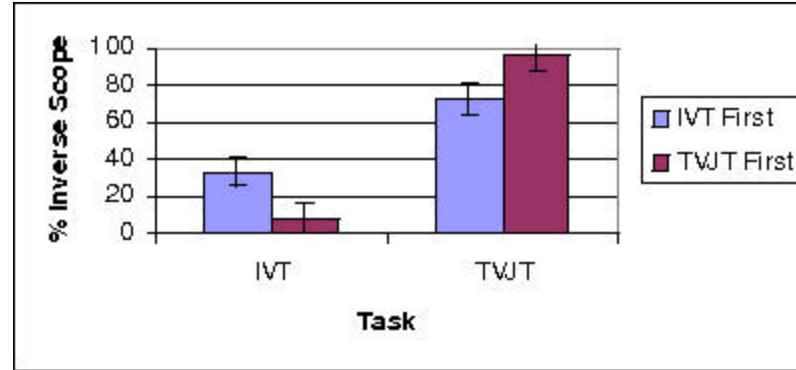


Figure 3.6: IVT: TVJT, graph of results

These percentages were entered into a mixed ANOVA with two factors: task (IVT vs. TVJT, within-subjects) $\times$ task-order (IVT-first vs. TVJT-first, between-subjects). We found a main effect of task ($F(1,19)=8.99$, $p=.007$), but not of task-order ($F(1,19)=.45$, $p=.50$), and a significant interaction between the two ($F(1,19)=7.19$, $p=.01$). Planned comparisons revealed a marginally significant difference between rates of inverse scope obtained in the TVJT between the two orders ($t(19)=1.87$, $p=.075$), but not in the IVT between orders ($t(19)=1.60$, $p=.124$).

Across all control trials on the IVT and filler trials on the TVJT, participants obtained the correct answer 100% of the time. Across all IVT targets, 1 trial was excluded from analysis due to failure to correspond to either interpretation. No TVJT trials were excluded from analysis.

Discussion

We found an effect of order on the rates of inverse scope obtained on the TVJT, but not on the IVT. It appears that obtaining surface scope interpretations on the IVT increased rates of surface scope interpretations on the TVJT. However, obtaining inverse scope interpretations on the TVJT did not lead to increased rates of inverse scope interpretations on the IVT.³ We conclude that there is a priming effect for surface scope interpretations in this experiment, but not for inverse scope interpretations.

Let us review our hypotheses. If it were the case that only non-initial interpretations were primable, then we predicted only the priming of inverse scope interpretations. This prediction was not borne out. The second hypothesis was that only dispreferred interpretations with respect to the preceding environment could be primed, which predicted priming only for surface scope interpretations. This prediction was borne out. We found surface scope priming when the IVT occurred first (when the preceding environment was the participant's inverse scope preference), but no priming for inverse scope interpretations when the TVJT came first. Crucially, we found no priming of the inverse scope interpretation when the preceding environment was composed of inverse scope interpretations (the adults' general preference). Therefore, it appears that priming only occurs when the prime differs in interpretation from the interpretations obtained in the preceding environment. This finding, although seemingly surprising, makes sense considering what we know about priming in other linguistic domains and in the context of our weather analogy: a temperature is likely to be surprising only when it differs greatly from your most recent experience.

³In fact, it appears to have lead to a marginal decrease of rates of inverse scope interpretations in the IVT. We have no explanation for this finding.

Let us summarize our findings to this point to recap how this hypothesis accounts for the results seen so far. Children, who obtain surface scope interpretations, show effects of the priming of inverse scope interpretations, but not surface scope interpretations. Adults, who generally prefer inverse scope interpretations, display the priming of surface scope interpretations, but not of inverse scope interpretations. In a task where adults’ preceding environment is set to surface scope interpretations (by the nature of the IVT), inverse scope interpretations can be primed. These results are summarized in Table 3.1.

Table 3.1: Summary of Priming Experiments
Children

Task	Preceding Environment	Prime	Result
TVJT	surface	<i>not every</i>	inverse scope priming
TVJT	surface	Partitive	inverse scope priming
TVJT	surface	Zero-morpheme	no surface scope priming

Adults

Task	Preceding Environment	Prime	Result
IVT	surface	<i>not every</i>	inverse scope priming
IVT	inverse	TVJT	no inverse scope priming
TVJT	inverse	IVT	surface scope priming

For each case, priming occurs when the prime is dispreferred with respect to the interpretations obtained in the preceding environment. These findings lend insight into the nature of LF priming, which could be obtained only by utilizing a novel task which allows adults to obtain surface scope interpretations.

3.2.2 Experiment Eleven: IVT: Revised Prime

Across all priming experiments, we observe a generalization that dispreferred interpretations can be primed, with respect to the interpretations obtained in the preceding environment, which is either the participant’s general preference, or an interpretation

obtained in a recent experiment. IVT: prime is a unique priming experiment, because it is the only experiment where the preceding environment differs from the participant's general preference. IVT: prime was designed so that the participant received 3 ambiguous target sentences, followed by 3 *not every* sentences and then 3 more ambiguous target sentences. That is, adults generally prefer inverse scope interpretations, but the first three trials of the experiment 'set' the preceding environment to a surface scope preference. We found effects of inverse scope priming in the last three trials, as compared to the first three trials. If it is the case that an adult's dispreferred interpretation is calculated (for priming) with respect to the interpretations obtained in the preceding environment, it may be the case that the first 3 ambiguous trials, in which surface scope interpretations obtain, are driving the priming effect.

In this experiment, we determine whether adults must experience IVT trials in order for the *not every* sentences to serve as a prime in IVT: prime. If our analysis is correct, then it must be the case that the first three trials are driving the priming effect. We do in fact find that the presence of the first three trials is responsible for the priming effect, suggesting that we can be confident in our formulation of 'preceding environment', and the observation that only interpretations that are dispreferred with respect to the preceding environment can be primed.

Materials and Methodology

The procedure and description of the task to participants remains identical to previous IVT experiments.

In this task, participants were presented with the same cards as IVT: specified domain. The first three target sentences are *not every* targets, as in (79). The second three targets are *every not* targets, as in (80).

(79) Not every princess has a crown

(80) Every frog doesn't have a fly

Unambiguous control trials remain the same as in IVT: specified domain.

All target and unambiguous control trials are scored identically to previous experiments.

Target and control items were mixed in a pseudo-random order. The order of presentation of items was counterbalanced, although the first three items were always of the form shown in (79). Two lists were created, with the second list presenting target and control items in the reverse order of the first list. Half of the participants received list one, and the other half received list two. 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.

Predictions

The analysis of priming that we have presented thus far predicts that priming effects only occur when the interpretation one obtains is different than interpretations that have been recently experienced. In general, one's experience is used for the calculation of the preceding environment. However, on the basis of results from IVT: prime, we have argued that the preceding environment is insufficient, and that interpretations obtained in a recent experiment can be used for the calculation. If this is the case (that the first three trials of IVT: prime are required to set the preceding environment to surface scope), then we predict that in this experiment, there will be no difference between the last three trials and the last three trials in IVT: specified domain (which contained no prime).

Alternatively, it may be the case that the inverse scope interpretations obtained in TVJT trials and *not every* trials are not the same, and as a result, may not behave in the same way as primes. *Not every* sentences unambiguously lead an adult to the logical form associated with inverse scope interpretations of *every x didn't* strings. The TVJT relies on the Principle of Charity, the assumption that an adult will assent to a true statement if available, for adults to select inverse scope interpretations. However, it is still the case that adults are selecting one interpretation of a scopally ambiguous string. It may be the case then, that *not every* sentences cause stronger priming effects than a TVJT. If this is the case, then it may be possible that guaranteed inverse scope interpretations obtained with *not every* sentences differ from the adults' interpretations obtained in the preceding environment (that is, general conversation). This hypothesis predicts that removing the first three ambiguous trials will have no effect, because the *not every* sentences still force one to obtain inverse scope interpretations, causing a priming effect.

Participants

20 undergraduate students at the University of Maryland, College Park, participated in this experiment, with 3 replaced in the design. Two participants were replaced because half of their answers did not correspond to either interpretation, and 1 for lifting cups before the completion of the target sentence. Participants received either monetary compensation or course credit for their participation. All participants were English native speakers over the age of 18.

Results

Across the last three target trials, participants obtained the inverse scope interpretation on 42% of the trials. This is not significantly different from the rate of inverse scope interpretations on the last three trials of IVT: specified domain ($t(38)=1.57$,

$p=.123$). One trial was excluded from analysis for not aligning with any interpretation. The appropriate response was obtained on 75% of *not every* trials. The majority of these errors (10/15) were on the first trial. Performance on unambiguous control items was 100% correct.

Discussion

The rate of inverse scope interpretations obtained in this experiment do not differ from the rate of inverse scope interpretations obtained in IVT: specified domain. Therefore, it appears when the first three trials were removed from IVT: prime, the *not every* sentences no longer created a priming effect.

These results further solidify our observation that priming occurs when a prime is a dispreferred interpretation. Furthermore, we found that the notion of ‘dispreferred’ is usually calculated with respect to a participant’s general preference, but that this preference can be overridden in an experimental situation. That is, the preceding environment must differ from the interpretation obtained by the prime in order to create priming effects.

3.2.3 Conclusions on Priming

We primarily are interested in the degree to which the IVT in adults mimics the behavior observed in children. Therefore, we investigated whether techniques that worked to increase inverse scope interpretations in children work the same way in adults. First, we found that the presence of a positive lead-in sentence worked in adults to increase inverse scope interpretations, just as in children. Second, we found that *not every* sentences prime inverse scope interpretations, as well. These two findings show that adults are aided by the same mechanisms as children, suggesting

that the IVT may be tapping into the same limiting capacities found generally in children.

The priming results raised additional questions concerning the nature of priming. Previously, we have been unable to test the relative priming of surface scope interpretations in adults because we have lacked the ability to obtain surface scope interpretations to serve as a prime. The IVT affords us a unique opportunity, because it reliably allows adults to obtain surface scope interpretations. By combining the IVT and the TVJT, we found that surface scope interpretations can be primed in adults. We conclude that LF priming mirrors lexical and s-structure priming in that LF priming occurs only with interpretations that are dispreferred, relative to the interpretations obtained in the preceding environment. In children, who prefer surface scope interpretations, we observe inverse scope priming, but not surface scope priming. In adults, in a task where adults have exposure to surface scope interpretations in the preceding environment (the IVT), we observe inverse scope priming.

This chapter lends motivation to studying children and adults in a unified way. Recall that our ultimate goal is to determine the nature of LF parsing, by investigating scopal ambiguity resolution in detail. In the next chapter, we will turn to investigating the processes underlying the interpretation of scopally ambiguous strings in children and adults.

Chapter 4

Interpreting Scope Ambiguity

In chapter 2, we confirmed that an advantage for surface scope interpretations holds in adults for sentences containing a universally quantified subject and negation. We also know that adults usually interpret these sentences with inverse scope interpretations in conversation (Musolino and Lidz, 2006). Together, these findings suggest that although there is an advantage for surface scope interpretations at one point in the comprehension process, this advantage can be overridden by the integration of discourse information. That is, in a situation where the inverse scope interpretation of a scopally ambiguous string is true, adults integrate the required discourse information and revise their interpretation to obtain the inverse scope interpretation.

In this chapter, we investigate children's interpretations of scopally ambiguous strings. Previous research provides mixed evidence concerning children's abilities to obtain inverse scope interpretations, with many studies finding that children show difficulty with accessing inverse scope interpretations at adult-like levels (Musolino et al., 2000; Lidz and Musolino, 2002; Musolino and Lidz, 2006). These findings are in contrast to those that find that children readily obtain inverse scope interpretations (Gualmini, 2004; Gualmini et al., 2005; Huksey et al., 2004). Because we are

attempting to identify the nature of the parser by viewing data from both children and adults, it is crucial to determine children’s abilities with respect to ambiguity resolution. We first suggest that previous experiments, that on the surface appear contradictory, actually present different timepoints of the same phenomenon. A review of this literature reveals that experiments test different age groups of children. We therefore perform an experiment testing children’s interpretations of scopally ambiguous strings across a wide age range. In this domain, we find that children exhibit U-shaped development, with 4 year olds obtaining inverse scope interpretations, 5 year olds obtaining surface scope interpretations, and adults obtaining inverse scope interpretations. This finding explains the previously conflicting findings of experiments testing children’s abilities to access inverse scope interpretations.

We suggest that this U-shaped development can be described by a two-stage maturation of children’s language processing capabilities. First, children acquire the adult-like advantage for surface scope interpretations. Second, children acquire the ability to revise their interpretation according to discourse information. To support this claim, we must show that children are unable to utilize situational information in the obtaining of an interpretation, but that adults do so. We present an experiment using both a sentence completion task and a TVJT, two tasks which present different experimental influences on interpretation. We find that although adults’ interpretations vary according to task, children’s do not. This finding is evidence that children’s interpretations are not affected by discourse information in the same way as adults’.

These findings allow us to further identify the differences between children and adults with respect to LF ambiguity resolution, and also to pinpoint the procedures that are utilized when adults process scopally ambiguous sentences.

4.1 Revisiting the Isomorphism Effect in Children

Given that children appear to differ from adults in this domain, we would like to know how the isomorphism effect can inform our understanding of the developing parser, the developing pragmatic system and the interactions between them. Although across many experiments, children fail to obtain inverse scope interpretations, this finding is not robust. Many experiments report that children can obtain inverse scope interpretations ((Musolino and Lidz, 2006), (Gualmini, 2004)). Previous research has identified both the presence and absence of an isomorphism effect across a wide variety of constructions and developmental timepoints. An assumption that is implicit in most previous studies is that the presence of non-adult-like behavior represents the starting point of the developmental trajectory. This, however, is an assumption subject to empirical investigation. In addition, in sentences lacking negation, children appear to behave like adults in their understanding of sentences involving quantification (Lidz et al., 2004). These findings provide strong evidence that children possess grammars that permit an adult-like understanding of sentences containing quantification. However, the exact source of children’s non-adult-like behavior is under debate. Some accounts propose that children’s hesitation to obtain inverse scope interpretations is due to non-adult-like parsers (Musolino and Lidz, 2006), while others claim that children’s surface scope interpretations are due to discourse factors (Gualmini, 2004; Hulsey et al., 2004).

Let us recall the two hypotheses regarding children’s shortcomings: the parsing hypothesis and the extra-linguistic hypothesis. The parsing hypothesis posits that children’s non-adult behavior can be attributed to an immature parser (Lidz and Musolino, 2002). For example, obtaining inverse scope interpretations requires revision. It may be the case the cost of revision prohibits children from obtaining inverse scope interpretations in their sentence processing. If the cost of revision is too high,

this would lead children to prefer surface scope interpretations of scopally ambiguous strings.

The extra-linguistic hypothesis posits that children's shortcomings derive from an inability to integrate discourse information in an adult-like way. Within this hypothesis, there are two important sub-hypotheses that we must articulate here: experimental infelicity and failure to utilize discourse information. One, it may be the case that children are unable to accommodate pragmatic infelicities (Gualmini, 2004). For example, in an experiment where the inverse scope interpretation is not made fully felicitous, children may lack the ability to overcome this infelicity (as adults can), and may be restricted to surface scope interpretations as a result. Two, it may be the case that children are unable to use discourse information in their environment as a means of guiding their choice of linguistic representation (Musolino and Lidz, 2006). For example, children may recognize the discourse function of an utterance, but may still fail to use this type of information as a trigger to perform revision of their initial linguistic analysis, thereby ultimately failing to integrate the discourse information into their language processing (Novick et al., 2005).

In this section, we want to determine the contributing factors to children's interpretations by distinguishing between parsing and discourse factors at the most general level. We present a re-investigation of the interpretation of scopal ambiguities by children. Previous research mainly tested 5 year olds, finding that these children displayed difficulty accessing inverse scope interpretations. This line of research inferred about younger children, suggesting that all children have preferences for the surface scope interpretation. In section 4.1.1, we show that although 5 year olds display an 'isomorphism effect', younger children (4 year olds) do not. This suggests that children's claimed preference for surface scope interpretations may not be as robust as

previously thought. These findings suggest that the parsing hypothesis, which claims that young children lack the parsing resources to obtain inverse scope interpretations, cannot be on the right track. Additionally, it cannot be the case that children’s behavior can be accounted for solely by experimental infelicity. We find a U-shaped pattern of development for the resolution of scope ambiguity. We argue that this developmental trajectory must be accounted for by a combination of the developing parser and discourse integration abilities, suggesting that previous accounts that appeal to only one source are insufficient. In section 4.2, we show that children lack the adult-like ability to alter their interpretation according to discourse information, lending support to the hypothesis that children’s discourse integration abilities are non-adult-like.

4.1.1 Experiment Twelve: TVJT Reinvestigation

The Observation of Isomorphism was identified by Musolino et al. (2000) for strings containing quantifiers and negation, as in (81) and (82).

(81) Every horse didn’t jump over the fence

(82) The detective didn’t find some guys

That is, children adhered to surface scope interpretations, while adults overwhelmingly obtained inverse scope interpretations in a TVJT. Gualmini (2004) ameliorated the isomorphism effect with the strings in (82), showing that under revised felicity conditions, children readily obtain inverse scope interpretations. In summary, it seems that previous research has identified both the presence and absence of an isomorphism effect across a wide variety of constructions and developmental timepoints, suggesting that the advantage for surface scope interpretations in children has a developmental trajectory that is poorly understood.

Gualmini shows that when the experimental situation is maximally felicitous, the advantage for surface scope interpretations is eliminated, for strings containing *some* and negation. Furthermore, Gualmini et al. (2005) showed that the effect can similarly be ameliorated for strings containing negation and a universal quantifier. Gualmini presented children with strings like (83), following a story where Caillou attempted to deliver all of the letters, but succeeded in only delivering some of them.

(83) Every letter wasn't delivered

Because the main focus of the story was whether or not all of the letters were delivered, the inverse scope interpretation of (83) was made felicitous. In this experiment, children (3;0-5;7, $m=4;8$) accepted the inverse scope interpretation 81% of the time. This finding suggests that the methodological changes affect children's interpretations. Therefore, the question remains as to whether manipulations in the experimental situation can entirely account for the advantage for surface scope interpretations, particularly with the construction we have been investigating, strings with a subject universal quantifier and negation.

However, while it may be the case that we can attribute the differences in rates of inverse scope interpretations between the Gualmini and Musolino et al. studies to methodological differences, methodology is not the only differing factor. The mean age of the children participating in the two experiments also differs. The average age of the children tested in the Gualmini experiment around four years old, a year younger than the five year olds tested in the Musolino et al. experiment. Therefore, we cannot determine if the differences in rates of inverse scope interpretations obtained across the studies are due to methodology or age differences.

Two hypotheses regarding the source of children's immaturity remain: the parsing hypothesis and the discourse hypothesis. We are interested in determining whether

the isomorphism effect, if it exists as reported in numerous studies, reflects the starting point of development or whether it occurs after a period of adult-like behavior. We test children across a wide age range on strings containing a subject universal quantifier and negation, a construction which has demonstrated a powerful isomorphism effect (Musolino et al., 2000). We find that although there exists a time during which children obtain mainly surface scope interpretations, this occurs after a period of adult-like behavior, suggesting that the isomorphism effect does not inform us about a child’s starting state. Any explanation for the advantage for surface scope interpretations in children must account both for why it occurs and also for why it shows a late onset. Any theory treating the effect as due to an inherent limitation of immature parsers or conversational agents must surely be incorrect. We argue that children have the ability to calculate inverse scope at a very early age, and the isomorphism effect occurs as a by-product of children’s developing adult-like parsers and discourse information integration abilities.

Materials and Methodology

We conducted a TVJT, modeled after the universal quantifier experiment in Musolino et al. (2000). Musolino et al. report results for 5 and 6 year olds, but Gualmini tested 4 year olds (although with a slightly different construction). Because we are attempting to determine the developmental trajectory of children’s interpretations, we require consistent data across the 4 and 5 year old age ranges with a single construction.

An example target sentence is shown in (84). The target sentence followed a story in which three cats talk about hiding behind a table, but ultimately attempt to hide behind the sofa. Two cats succeed in hiding behind the sofa, and one stays behind

the box. The final scene is shown in Figure (4.1).¹

(84) Every cat didn't hide behind the sofa



Figure 4.1: TVJT Reinvestigation: final scene

The TVJT was conducted on a computer screen, with animated characters. The movement of each character was controlled by the experimenter, who told the story verbally. The target sentence was recorded. To the best of our knowledge, these stories mimic tasks done with act-out toys, but have the advantage of requiring only one experimenter.

Each child received 2 warm-up items, 6 target items and two filler items. The filler sentences were dynamic, so that the experimenter could choose either a true or false sentence, depending on the child's previous responses. This ensured that the child's responses varied across the experiment.

Children were tested one at a time in a quiet testing room away from other children. The child was introduced to the computer, and told how to play the game. The

¹The TVJT was conducted on a computer screen, with animated characters. These stories mimic tasks done with act-out toys, but have the advantage of requiring only one experimenter. A within-subjects experiment comparing these factors is presented in section 4.1.2.

puppet, a dog, was also on the computer screen, and gave an introduction before beginning the experiment. The target and filler items were mixed in a pseudo-random order. The entire session lasted less than 20 minutes.

Predictions

There are two hypotheses concerning the description of children's behaviors: the parsing hypothesis and the extra-linguistic hypothesis.

One hypothesis is that children, either due to cognitive limitations or the structure of their parser, are unable to access the inverse scope interpretation except in cases with significant discourse assistance (Lidz and Musolino, 2002). One implementation of this idea is that children have great difficulty performing the revision required to access the inverse scope interpretation. According this hypothesis, there is a genuine isomorphism effect, deriving from the difficulty required in calculating inverse scope. This hypothesis predicts difficulty with inverse scope at the earliest possible age, persisting until the child is old enough to have developed adult-like parsing resources. According to this hypothesis, there need not be any shortcoming in the child's discourse integration abilities.

A second hypothesis is that children's parsers are completely adult-like, and the isomorphism effect derives completely from children's inability to accommodate infelicitous experimental conditions (Gualmini, 2004). Accommodation is a specific instance of failure to integrate discourse information in an adult-like way. According to this hypothesis, given fair experimental conditions, children will obtain inverse scope at a very early age, and we expect to find no isomorphism effect across children.

A third hypothesis is that the previous results can be accounted for by an age effect. That is, four year olds readily obtain inverse scope interpretations, but five year olds do not. This suggests we should find a U-shaped curve in the current experiment.

Participants

30 children participated in this experiment. Children were recruited from a preschool at the University of Maryland. The children were divided into two groups, four year olds and five year olds, each group consisting of 15 children. One child was replaced in the design for incorrectly answering both filler items. The four year old group ranged from 4;5-5;2 (mean 4;9) and the five year old group aged 5;3-5;7 (mean 5;4). The ages of these two groups are statistically different ($p < .01$). These age groups were chosen to maximally match the previous studies conducted. Gualmini (2004) tested children younger than 5;0, while Musolino and Lidz (2006) tested children older than 5;0. Furthermore, Musolino et al. (2000) found an age difference in their study around the age of 5;2. Therefore, we selected 5;2 as a separation point, although we make no claims that critical changes occur at this exact age. Further research must be done to determine the causes underlying these observed age differences.

12 adults from the University of Maryland participated in this experiment as a control. Participants received either monetary compensation or course credit for their participation. All participants were English native speakers over the age of 18.

Results

Adult participants accepted the target sentences (consistent with inverse scope interpretations) 76% of the time. This acceptance rate is lower than previous studies (Musolino et al., 2000; Musolino and Lidz, 2006), including our own IVT: TVJT, and we have no explanation for this finding. Performance with fillers across all groups

is over 90%, suggesting overall good performance on the task. No target trials were excluded from analysis.

We found that the four year olds accepted the inverse scope interpretation of the target sentences 81% of the time, at the same rate as adults. All children were asked for justification for their responses, and most children gave justifications consistent with the their true/false answers, suggesting their acceptance of the puppet's statement reflects obtaining the inverse scope interpretation.

The five year olds accepted the inverse scope interpretation 44% of the time. This rate is significantly lower than the rate shown by the four year olds ($t(28)=2.05$, $p < .05$), and marginally significantly different from the adults ($t(25)=1.98$, $p < .06$). The five year old group contains 7 children who never access the inverse scope interpretation, a bimodal distribution suggesting that these children fail to access the inverse scope interpretation.

Discussion

Let's recall the hypotheses and predicted results. If it is the case that children have non-adult parsers, we expect younger children to behave just like the five year olds, in being restricted to the surface scope interpretation of the target sentence. However, the four year olds accessed inverse scope at the same rates as adults, suggesting this cannot be the case. Younger children can obtain the inverse scope interpretation, and presumably younger children do not have more parsing resources than older children, so we conclude that the isomorphism effect in five year olds cannot be due to immaturity of the sentence parser, as claimed in Lidz and Musolino (2002).

If it is the case that the advantage for surface scope interpretations is a result of infelicitous materials, then we too, expect similar rates of acceptance to persist into the four year olds. Although previous research has shown that there is certainly an effect of felicity on interpretation, in this experiment, we see interpretation differences across age, and not as a product of methodology. Therefore, we conclude that the isomorphism effect does not solely derive from a failure to experimentally meet felicity conditions.

If it is the case that children's age accounts for interpretation, then we expected a U-shaped developmental curve to account for previous results. It appears that this is the case. Musolino et al. and Musolino and Lidz report low rates of acceptance of the inverse scope interpretation, with data from 5 year olds. We replicated this 'isomorphism effect' with 5 year olds with the universal quantifier in subject position. However, we found that the isomorphism effect holds only in a limited window in language development, and does not represent the starting point of acquisition. We found that four year olds are adult-like in obtaining inverse scope interpretations, replicating the Gualmini results.

Typically, a U-shaped curve of development is analyzed as the child developing adult-like mechanisms in two stages. Let us look at a classic example, the acquisition of the English past tense. We will outline the analysis, and then suggest that the child's acquisition of adult-like scope interpretive abilities shows a similar pattern of development. For the past tense forms of irregular verbs in English, children demonstrate a U-shaped curve of development. That is, they begin by using the proper morphological form (such as *went* for the verb *to go*), then begin using a regularized past tense form (such as *goed*), then finally utilize the appropriate form *went* (Brown, 1972). At first glance, this pattern seems perplexing, because the child appears to

become non-adult-like after a period of producing adult-like forms. However, this pattern is easily explained by appealing to the child's developing abilities to form the past tense in an adult-like way. Children in the youngest stages, although they produce adult-like forms, are obtaining this form by non-adult-like means. That is, they are simply mimicking their input. At the second stage, children have acquired a general rule for forming the past tense, but have not learned that some verbs are irregular. At the final stage, the child has not only the general rule, but also has learned that some verbs form exceptions to this rule. This is depicted in Figure (4.2).

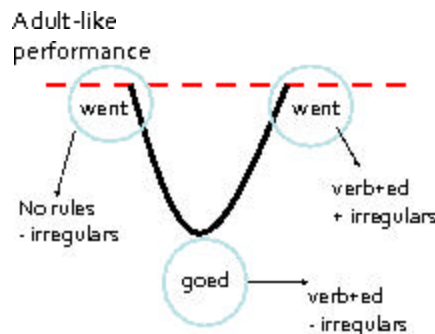


Figure 4.2: Acquisition of the English past tense

Notice that the child, to obtain adult competence, must acquire two things. First, they must acquire the general rule for forming the past tense, *verb + ed*. Second, they must acquire an exception that overrides the implementation of this general rule for irregular verbs (like *to go*). The crucial part of this analysis is that the apparent non-adult-like behavior of the child represents the child developing the mechanisms that underlie adult behavior. We can restate the U-shape curve in terms of three stages of development: 1) apparent adult-like behavior by non-adult-like means, 2) non-adult-like behavior by using adult-like processes, and 3) fully adult-like behavior.

We suggest that the same type of account can describe children's development of scope ambiguity resolution. Recall that we have observed that children initially obtain inverse scope interpretations (which appears adult-like), then go through a period of

obtaining surface scope interpretations. Also recall that when an adult obtains an inverse scope interpretation, two steps are involved. Depending on the proper analysis of adults, adults either have unbiased parsers (under the extra-linguistic hypothesis), or initially select surface scope interpretations (under the parsing hypothesis). In either case, adults can access an inverse scope interpretation by using some amount of discourse information. This is quite similar to the two steps required in accounting for an adult-like description of the English past tense, which requires a general rule, but also the ability to allow exceptions to this rule in the case of irregular verbs. With respect to scope, we propose that the U-shaped curve observed can be described by children first acquiring an adult-like parser (that is, either an unbiased parser or a surface-scope-first parser), and then adult-like revision capabilities (this process will be detailed more explicitly in section (5.1)). We claim that 4 year olds are not obtaining inverse scope interpretations in an adult-like way, but are simply mimicking the inverse scope interpretations observed in the input. Notice, it is not the case that children lack the parsing ability to obtain an interpretation, but simply have a parsing preference that adults do not. Next, the child obtains the adult-like parser, which depending on the formulation of the adult system, will either output solely surface scope interpretations or interpretations at chance. However, at this stage, the child is lacking the ability to revise their interpretation according to discourse information, so remains with their initial interpretation. Finally, the child acquires the ability to utilize discourse and situational information as a way of revising their interpretation. This is depicted in Figure (4.3).

This description of U-shaped development can also be paraphrased into three stages: 1) apparent adult-like behavior by non-adult-like means, 2) non-adult-like behavior by using adult-like processes, and 3) fully adult-like behavior. In the upcoming chapters, we will make the description of the child system more explicit as we further develop our description of the adult LF parser.

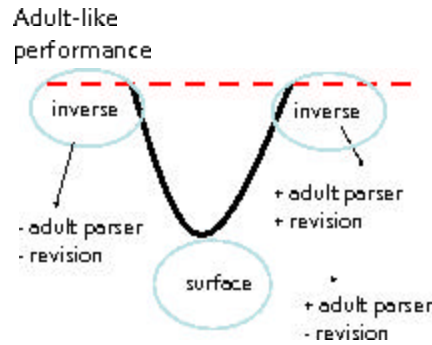


Figure 4.3: Acquisition of Scope Ambiguity Resolution

This analysis yields an interesting explanation for the isomorphism effect observed with another construction. Lidz and Musolino (2002) tested four year olds with strings containing negation and a numerically quantified object, as in (85).

(85) The detective didn't find two guys

(85) is ambiguous, meaning either *it is not the case the detective found two guys* (the surface scope interpretation) or *there are two guys the detective failed to find* (the inverse scope interpretation). Musolino and Lidz tested 12 children (3;11-4;11, $m=4;4$) in a situation where the inverse scope interpretation was true. Children accepted the statement in (85) as a description of the scene only 33% of the time, in contrast to the adults' 93% acceptance rate. These results suggest that with respect to these strings, children prefer surface scope interpretations. Because this experiment tested four year old children, at first glance, these results appear to contradict our proposed model. That is, it is not universally the case that four year olds adhere to inverse scope interpretations. However, our claim is that four year olds prefer the interpretation that appears most frequently in their input. Indeed, Musolino and Lidz find that adults prefer the surface scope interpretation of (85). In a situation where both interpretations of (85) are true, adults consent, and give justifications that support the surface scope interpretation. In the case where the surface scope interpretation is preferred for adults, our U-shaped model would suggest that both

four and five year olds would adhere to surface scope interpretations. Therefore, these results provide further support for our analysis.

These results make a prediction concerning inverse scope interpretations in four year olds for strings containing a universally quantified subject and negation, that is unfortunately, difficult to test. Given that it is the case that 4 year olds have a parser bias for inverse scope interpretations, and have difficulty using discourse information as a means of revising interpretation, we would predict that these children would be stuck with inverse scope interpretations. That is, in a TVJT where no horses jump over the fence, children in this age range should judge a sentence like (86) as false.

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

That is, if children in this age range think (86) means *not all of the horses jumped over the fence*, and furthermore compute the implicature that *not all* means *some*, then they should judge this sentence to be false. The difficulty in testing this prediction is that if it is the case that children of this age consent to a sentence like (86) in a situation where none of the horses jumped over the fence, we cannot be certain if they are accessing the surface scope interpretation, or if they are failing to compute the relevant implicature (this is particularly likely, as it has been shown that children have difficulty computing implicatures (Noveck, 2001; Musolino and Lidz, 2006)). Therefore, we will need to discover other methods to determine the extent to which 4 year olds are able to obtain surface scope interpretations.

We must show that children lack the ability to integrate discourse information into final interpretation, in order to be confident in our analysis of the development of scope ambiguity resolution. We turn to an experiment to determine whether this is the case in section 4.2. However, first, we present a second experiment that confirms that the results presented in this section are robust, and cannot be described by the

novel experimental methodology used.

4.1.2 Experiment Thirteen: Presentation Comparison

We presented results from an experiment that found that four year olds readily obtain inverse scope readings. This result is surprising in light of previous findings. However, there is a methodological difference between previous experiments and our own that we must consider as a potential contributor to interpretive differences. While previous experiments used only toy characters, we performed the TVJT on a computer screen. Therefore, we must verify that there is no difference in the interpretations obtained between computerized and toy TVJTs.² We find no difference between the two modes of presentation, suggesting that we can be confident in our previous results.

Materials and Methodology

This experiment contained 6 target TVJT stories, modeled after Musolino et al. (2000). For each story, a computerized version and a toy version was created. An example target sentence is shown in (87).

(87) Every giraffe didn't pick an apple

The final scene of the computerized story that precedes (87) is shown in Figure (4.4).

A photograph of the final scene of the toy version is shown in (4.5).

A description of the story that accompanies each of these final scenes is below.

Three giraffes were playing outside, and consider going home, but decide that they are hungry. They all decide to pick an apple. The first two

²Thanks to Kristin Koch, undergraduate research assistant, for the writing of stories and implementing of this experiment.

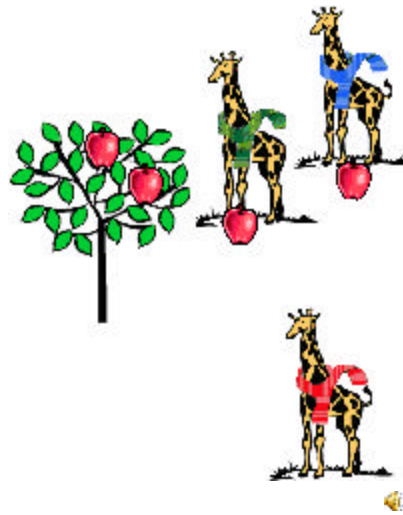


Figure 4.4: Presentation Comparison: Final computerized scene

giraffes succeed, but the only apples that are left are too high for the third giraffe. He does not pick an apple in the end.

For each story, the toy and computerized versions were developed simultaneously to ensure the closest possible fit between the two modes of presentation. As in TVJT reinvestigation, the computerized stories were animated using Powerpoint, so that the stories mimic toy stories as closely as possible.

Each child saw 3 computerized targets, and 3 toy targets. Target and filler items were mixed in a pseudo-random order. Order and task-story combinations were counterbalanced. Four lists were created. Target and filler items were divided into two groups: Group A and Group B. List one and two presented Group A first, and lists three and four presented Group B first. In lists one and three, the first group presented was shown in computerized format, in lists two and four, the first group presented was shown in live toy format. Participants were divided evenly amongst the



Figure 4.5: Presentation Comparison: Final toy scene

lists. Each child received two warm up stories,³ and two filler trials. The fillers were dynamic, so that the experimenter could choose either a true or false target sentence, depending on the child's previous responses. This ensured that the child's responses varied across the experiment.

Children were tested individually in a quiet testing room. The entire session lasted approximately 15 minutes.

Predictions

If it is the case that the mode of presentation affects children's interpretations, then we expect to observe differences across tasks. Specifically, to accord with previous results, we would expect higher rates of surface scope interpretations on toy trials than on computerized trials.

If mode of presentation does not influence the interpretations obtained by children, then we expect no effect of condition on interpretation. The finding that children's

³The warm-up stories followed the format of the first target presented to the child.

responses do not change as a result of the computerized presentation would strengthen our confidence in the results of the previous experiment.

Participants

We tested 16 children from preschools surrounding the University of Maryland, College Park. Two children were replaced in the design for incorrectly answering both filler items (1) or failure to complete the experiment (1).

Children ranged in age from 4;2-5;10, with a mean age of 5;2.

Results

Across all computerized trials, children obtained inverse scope interpretations on 70.1% of trials. Across all toy trials, children obtained inverse scope interpretations on 75% of trials. These rates are not significantly different ($t(15)=-1.0$, $p=.33$). Therefore, we fail to find a difference between conditions. Individual children also reveal no difference between conditions. For each child, we can calculate a difference score between story types, that is # inverse scope interpretations on computerized trials minus # inverse scope interpretations on toy trials. Of the 16 children, 12 children obtained a difference score of zero. Of the remaining 4 children, 2 children obtained difference score of (-1), and 1 obtained a score of (1). Therefore, we conclude that there is no effect of task on interpretation.

Just as in TVJT reinvestigation, we can divide children into either 4 year olds or 5 year olds. The four year old group (4;2-4;11, mean= 4;6) contains 6 children, and the 5 year old group (5;2-5;10, mean= 5;6) contains 10 children.⁴ The four year olds obtained inverse scope in 91.6% of trials. This group contained 5 children who

⁴These age groups are unequal because age was not a predetermined factor for this experiment.

obtained inverse scope on all 6 trials, and one who obtained inverse scope on half (and these were divided between computerized and toy trials). The five year olds obtained inverse scope interpretations on 61.6% of trials, with 4 children obtaining inverse scope on all 6 trials. Three of the 5 year olds obtained surface scope interpretations on all trials. These effects approach significance ($t(13)=2.16$, $p=.06$) and mirror the previously obtained age effects.

Discussion

We can conclude two things from this experiment. One, with children, the differences between the materials used in TVJT reinvestigation and the materials used in previous experiments yield no interpretive effect.⁵ The results from this experiment show that it cannot be the case that our results from TVJT reinvestigation are due to children obtaining a higher rate of inverse scope interpretations when the targets are presented on the computer.

Second, we replicated a trend of age effects. This suggests that, although we have confirmed a tendency for inverse scope interpretations to decrease at one developmental timepoint, that age is not the correct correlate. That is, we can reliably replicate 4 year olds obtaining high rates of inverse scope interpretations, and find a number of 5 year olds obtaining only surface scope interpretations. However, interpretation differences likely correlate with a developmental measure, of which age is only a weak correlate (such as either psychological or linguistic development). It remains to be seen what measure obtains the proper predictive power.

⁵In TVJT reinvestigation, we found that our adults obtain fewer rates of inverse scope on a computerized presentation than with a toy presentation, although this has not been tested within subjects. We have no explanation for this finding.

4.1.3 Conclusions on Children’s Abilities

In this section, we presented two findings. One, we found that young children readily obtain inverse scope interpretations. This finding suggests that a simple competence hypothesis (such as immaturity of the parser) is insufficient for capturing the isomorphism effect. Two, we found that around the age of five years old, children go through a period of obtaining high rates of surface scope interpretations.

Our previous experiments have suggested that adult behavior can be described by an initial adherence to surface scope interpretations, with revision to inverse scope interpretations in the proper situation. We have proposed in this section that children’s U-shaped development of scope ambiguity resolution can be described by children’s developing adherence to these two features of adult mechanisms. We propose that young children obtain inverse scope interpretations by non-adult-like means, perhaps simply mimicking their input. Children then acquire adult-like parsers. However, children at this stage lack the ability to utilize discourse information to revise their interpretation, and remain stuck with surface scope interpretations. Finally, children acquire the ability to make use of discourse information to influence their choice of interpretation.

This account makes a strong prediction: that both four and five year olds are unable to make use of discourse information to influence their choice of interpretation. We investigate this issue in section 4.2.2. We show that children’s interpretations do not vary according to experimental environment, supporting our proposed explanation of U-shaped development. We also want to know what the development of scope interpretations reveals about the adult parser. Recall that to this point, we have yet to determine the source of the advantage for the surface scope interpretation in adults. We investigate a model in section 5.1, which suggests that these data points

cannot be captured by an adult parser with a bias for surface scope interpretations, suggesting this advantage must be captured external to parsing preferences.

4.2 Discourse in Ambiguity Resolution

In this section, we investigate the effect of discourse information on interpretation in children and adults. Our results with four year olds suggests that they are ‘adult-like’ in obtaining inverse scope interpretations. However, we must hesitate to take results from TVJTs as significant findings about preferences, as they were designed only to test the availability of an interpretation. We perform a TVJT and sentence completion task with both children and adults. We find that children are not adult-like across experimental tasks. This suggests that, even though young children display some knowledge of adult-like performance, they have not fully obtained adult interpretive abilities.

This finding brings us closer to fully identifying the contributing factors to attaching a logical form to a string in both children and adults.

4.2.1 Experiment Fourteen: Adult TVJT-Sentence Completion

In this section, we present an investigation into task differences with respect to interpretation of scope ambiguity in adults. It has been reported that adults, who adhere to inverse scope interpretations in a Truth Value Judgment Task (TVJT), also demonstrate a preference for inverse scope interpretations in comprehension (Musolino et al., 2000). However, this observation is simply a correlation, and was not meant to imply a deep connection. We show that this preference for inverse scope is not consistent across all tasks, suggesting that there may be no overall preference for inverse scope

interpretations in adults. Furthermore, we suggest that fluctuation of interpretation across tasks is a hallmark of adult-like performance.

Materials and Methodology

As in the forced choice task, participants are presented with multiple items that can be used to complete a scopally ambiguous string, as in (88).

(88) Every dog didn't decorate the fire hydrant that's...

The preceding story makes available a fire hydrant that *none of the dogs decorated* (the blue one) and a fire hydrant that *not all of the dogs decorated* (the yellow one), as shown below.⁶

Here, there is a purple, red and green dog, with their lollipop decorations. The fireman uses a hose as a decoration. It doesn't look like any of the dogs decorated the blue fire hydrant, but the fireman did. And it looks like the purple and green dogs decorated the yellow fire hydrant, but not the red dog.

Choosing *the blue fire hydrant* to complete the sentence is consistent with a surface scope interpretation of (88), while selecting *the yellow fire hydrant* is consistent with the inverse scope interpretation.

In this experiment, participants participated in two conditions: a sentence completion task and a Truth-Value Judgment Task. For both tasks, subjects are presented with a picture on the computer screen, and the accompanying story.

⁶The stimuli used for this task are identical to the stimuli used for the non-speeded condition of the forced choice task.



Figure 4.6: Adult TVJT-Sentence Completion: Target presentation slide

Each condition of the experiment contained 6 target items, with warm-ups and 3 filler items. The filler sentences contain a universal quantifier and no negation, as in (89).⁷

(89) Every chef glazed the flower pot

The filler items were presented in the same manner as the target items, but with modifications to the story. An example story is below, and with the story depicted in Figure (4.7).

Here, there is a yellow, blue and red chef, with their glaze. Kermit uses green glaze. It looks like the yellow, blue and red chefs glazed the pot made for cooking. And it looks like the yellow and blue chefs glazed the flower pot, but not the red chef.

The experimental items were divided into two lists. Each subject received one list in the form of a sentence completion task, and the other list in a TVJT, each of which are described below. Adults received the sentence completion task first, followed by the TVJT. Half of the participants received list one for the sentence completion task,

⁷The format of the sentence is all that changes across tasks.



Figure 4.7: Adult TVJT-Sentence Completion: Filler presentation Slide

while the other half received list two. Participants received the alternate list for the TVJT.

Sentence Completion Task: Adults

We conducted a sentence completion task, in which both interpretations of the scopally ambiguous target sentence are true and present.

The participant is told that a puppet attempts to say what happens in the scenes presented, but is not very good with his colors. The participant's task is to help him complete a sentence that is an accurate statement about what happened in the story. The puppet, after watching the story presented, says the prompt in (90).

(90) Every dog didn't decorate the fire hydrant that's... umm...

The participant's task is to complete the sentence. If participants obtain an inverse scope interpretation, they will answer *yellow*, but if they access the surface scope interpretation, they will answer *blue*. As in the speeded forced choice task, the color answer is indicative of interpretation that is obtained. Participants are asked to

choose the best answer in the case that more than one is available to them.⁸

TVJT: Adults

In the TVJT version, the participants received the same presentation of the story, but were told that their task was to tell the puppet if his statement in (91) was right or wrong.

(91) Every dog didn't decorate the yellow fire hydrant

In the target sentence, the puppet utters a statement that is true on the inverse scope interpretation. Recall, the item associated with the inverse scope interpretation is counterbalanced across items according to which side it appears on, so that participants do not receive any hints by relying on one particular side of the screen.

In this set of experiments, the participant received the sentence completion task first. If it is the case that TVJTs are capable of pushing an adult toward the true interpretation, then this may cause priming effects if this task were to come first. For this reason, the TVJT was always run after the sentence completion experiment. A distractor task appeared between the two conditions, so that they appeared to be different experiments.

The two tasks combined lasted approximately 20 minutes.

Predictions

We predict high rates of acceptance of the inverse scope interpretation in the TVJT condition of the experiment.

⁸Participants are told this only if they give a response saying that both are acceptable. Otherwise, participants are not told this, in order to avoid the raising of the idea that the sentences may be ambiguous.

We have two hypotheses regarding performance on the sentence completion task. If it is the case that participants, in this task, behave identically to the non-speeded condition of the forced choice task, then we predict around 50% acceptance of the inverse scope interpretation. In this case, we should expect a difference between the rates of inverse scope interpretations between the two tasks.

Alternatively, it may be the case that adults generally prefer inverse scope interpretations, which would result in high rates of obtaining inverse scope in the sentence completion task. In this case, it would be mysterious why adults were at chance performance in the non-speeded condition of the forced choice task. This hypothesis predicts similar high rates of inverse scope interpretations across tasks.

Participants

20 adult undergraduates from the University of Maryland participated in this experiment. None were excluded from analysis. Participants received either monetary compensation or course credit for their participation. All participants were English native speakers over the age of 18.

Results

Across all trials in the sentence completion condition, adults gave answers consistent with the inverse scope interpretation 40% of the time. This result is not significantly different from the results obtained in the non-speeded condition of the forced choice task ($t(38)=2.02$, $p=.44$). No trials were excluded from analysis. For 3 trials, participants responded that the sentence could be completed by either answer. We have coded these responses as obtaining the inverse scope interpretation. If these are coded as surface scope responses, the rate of inverse scope interpretations is 37.5%. If these trials are excluded from analysis, the rate of inverse scope interpretations is 38.4%.

For the remaining statistics, we code these responses as inverse scope responses, because the inverse scope interpretation was obtained.

Across all trials in the TVJT, 74.1% accepted the statement, obtaining the inverse scope interpretation. This is the same percentage we have found in other computerized TVJTs. This is significantly different from the sentence completion task ($t(19)=4.19$, $p<.001$). No trials were excluded from analysis. One trial was reported as ambiguous.

Adults responded correctly to 100% of the filler trials in the TVJT, and 96% of trials in the sentence completion task.

Discussion

First, these results confirm previous findings that in a TVJT, adults obtain largely inverse scope interpretations.

We find that in a sentence completion task, adults do not demonstrate a preference for inverse scope. This suggests two things. One, it suggests that the preference shown for inverse scope in a TVJT is a result of a task effect. For example, it is likely that adults's adherence to inverse scope interpretations is a result of the Principle of Charity, the desire to consent to a true statement, as opposed to a genuine preference. Second, these results suggest that, in a task with little constraints imposed by the discourse or situation, adults are about at chance with respect to what interpretation they obtain.

In summary, we find an effect of task on adults' interpretations. Therefore, obtaining inverse scope interpretations on a TVJT is not a sufficient measure of 'adult-like performance'. True adult performance integrates discourse information into final interpretation, yielding interpretations that differ as a function of the discourse context.

If it is the case that 4 year olds have fully adult sentence processing capacities, we expect a similar pattern of performance.

4.2.2 Experiment Fifteen: Pilot Child TVJT-Sentence Completion

We have observed that four year olds are able to obtain inverse scope interpretations in a TVJT. We have also observed a difference in interpretations adults obtain in a TVJT and a sentence completion task. This finding suggests that adult-like interpretation is dependent, at least partially, on task-specific discourse information. Therefore, if it is the case that 4 year olds are adult-like, we expect to find similar fluctuation of interpretation across tasks.

Materials and Methodology

For this experiment, we used the TVJT task (materials from TVJT reinvestigation) and one list from the sentence completion condition of Adult TVJT-Sentence Completion.

We tested children on the TVJT, described in section 4.1.1. If a child obtained 5 or more (out of 6) inverse scope interpretations on the TVJT, we tested them on the sentence completion task, because we are interested if children who obtain inverse scope interpretations on a TVJT obtain surface scope interpretations on a sentence completion task.

Children were introduced to the sentence completion experiment in the same way as adults. That is, children are shown a puppet, and told that the puppet was not very good at describing what happened in the story, and that they would need to

help the puppet by completing his sentences.

Children were tested individually in a quiet testing room. The entire session lasted approximately 15 minutes.

Predictions

If it is the case that four year olds are completely adult-like, then we would expect their rates of inverse scope interpretations to drop in a sentence completion task, as compared to the number of inverse scope interpretations obtained in a TVJT.

If it is the case that our analysis of the U-shaped development is on the right track, then four year olds are not accessing the inverse scope interpretation by revising through the surface scope interpretation. Furthermore, it must be the case that children are not making use of discourse information as a means of guiding their choice of final interpretation, as suggested by Trueswell et al. (1999). If this is the case, we expect children's interpretations to remain largely the same across tasks. That is, the different discourse environments available across tasks will not have any influence on the child's interpretation.

Participants

12 children (4;8-5;5, mean=4;11), who obtained largely inverse scope interpretations on the TVJT on the sentence completion task participated in this experiment. No children were excluded. This experiment was conducted approximately 2 weeks after the conclusion of TVJT reinvestigation, the experiment for which the child's TVJT rates were obtained.

Results

Across all children, the rate of inverse scope completions is 76%. Recall, because we only selected children who obtained high rates of inverse scope interpretations on the TVJT, their rates of inverse scope interpretations on that task was 94.4%. This difference approaches significance ($t(11)=1.99$, $p=.07$). No trials were excluded from analysis.

For each child, we can look at how their performance on this task differed from number of inverse scope interpretations on the TVJT. For 8 of the 12 kids, the number of inverse scope interpretations on the sentence completion task and TVJT differed by one or fewer responses. Three of the four remaining children had scores that differed by more than 3. For these three children, they obtained largely inverse scope interpretations on the TVJT, but half or more surface scope interpretations on the sentence completion. Two of these children were aged 4;11, and one was aged 5;1.

Across all filler items, performance was 97.8% accurate.

Discussion

From the results of this experiment, we cannot conclusively determine if children's interpretations differ according to task. While the rates of inverse scope obtained approach significance across the two tasks, only 4 children (1/3 of the total) have rates of obtaining inverse scope interpretations that drop by more than one response in a sentence completion task. It appears that the effect of the entire group is carried by the few number of children whose performance does vary according to task.

Therefore, we must perform the same experiment with a wide range of children, to determine the existence of an effect.

4.2.3 Experiment Sixteen: Child TVJT-Sentence Completion

Although previous tests of scope ambiguity resolution have relied on TVJTs, these tasks, alone are not sufficient to characterize adult-like performance. We have showed that although adults obtain high rates of inverse scope interpretations on a TVJT, they access the two interpretations equally in a sentence completion task. This suggests that a characterization of the adult scope ambiguity resolution system requires a component that can alter interpretation in order to meet discourse (and therefore, task) demands.

Independently, children's abilities to use task information to disambiguate ambiguity in an adult-like way has been called into question (Trueswell et al., 1999; Musolino et al., 2000). Therefore, although we observe young children obtaining inverse scope interpretations, this is not sufficient evidence to conclude that children have adult-like systems for resolving scope ambiguity. To be adult-like requires not just obtaining inverse scope interpretations in certain tasks, but also the ability to alter interpretation by situation on the basis of discourse information.

The results of our previous experiment were inconclusive. Therefore, we conduct a comparison of TVJT and sentence completion tasks, identical to Adult TVJT-sentence completion, within children. We find that children's interpretations do not vary according to task, suggesting that children's discourse integration abilities are not adult-like. This finding strengthens our confidence that our analysis for the U-shaped behavior observed in children is on the right track.

Materials and Methodology

In this experiment, each child received 6 TVJT target trials, and 6 sentence completion target trials, using the same materials as Adult TVJT-sentence completion (including an identical number of warm-up and filler trials). For the children, order between conditions was counterbalanced. Four lists were created. Target and filler items were divided into two groups: Group A and Group B. List one and two presented Group A first, and lists three and four presented Group B first. In lists one and three, the first group presented was the TVJT, in lists two and four, the first group presented was the sentence completion task.

Children were tested individually in a quiet testing room. If children had difficulty sitting for both parts of the experiment, they were asked to return for the second part on a separate day. Therefore, the children frequently performed the conditions on different days, with each section lasting approximately 15 minutes.⁹

Predictions

If it is the case that children are truly adult-like, then we expect that in this experiment, we will observe a difference between rates of inverse scope interpretations obtained in the two different conditions. Specifically, we expect higher rates of surface scope interpretations in the TVJT trials than in the sentence completion trials.

Alternatively, it may be the case that children are arriving at interpretations via non-adult-like means. In this case, we expect similar rates of inverse scope interpretations obtained across tasks. Recall that our analysis of U-shaped development of ambiguity resolution suggests that neither four nor five year olds have the ability to

⁹It is common for child experiments to be split over different days, and we do not believe this difference from adults to have any meaningful impact on the data.

utilize discourse information in accessing an interpretation. This suggests that all children should access the same interpretation, regardless of task information.

Participants

We tested 20 children from preschools in the College Park, Maryland area. Four children were replaced in the design for incorrectly answering 50% of the filler items. Children ranged in age from 4;1 - 5;11 (mean=5;0).

Results

Across all trials in the sentence completion condition, children gave answers consistent with the inverse scope interpretation 42% of the time. No trials were excluded from analysis.

Across all trials in the TVJT, children accepted the statement 45% of the time, obtaining the inverse scope interpretation. No trials were excluded from analysis.

Children responded correctly to 88.3% of the filler trials in the TVJT, and 81.7% of trials in the sentence completion task.

The results of child and adult TVJT-sentence completion, combined, are shown in Figure (4.8).

These percentages were entered into a mixed ANOVA with two factors: task (sentence completion vs. TVJT, within-subjects) $\times$ age (adult vs. child, between-subjects). We found a main effect of task ($F(1,38)=9.7$, $p=.003$), but not of age ($F(1,38)=1.4$, $p=.23$), and a significant interaction between the two ($F(1,38)=6.5$, $p=.01$). Planned comparisons revealed a significant difference between child and adult acceptance rates in the TVJT ($t(38)=2.25$, $p=.03$), and a significant difference be-

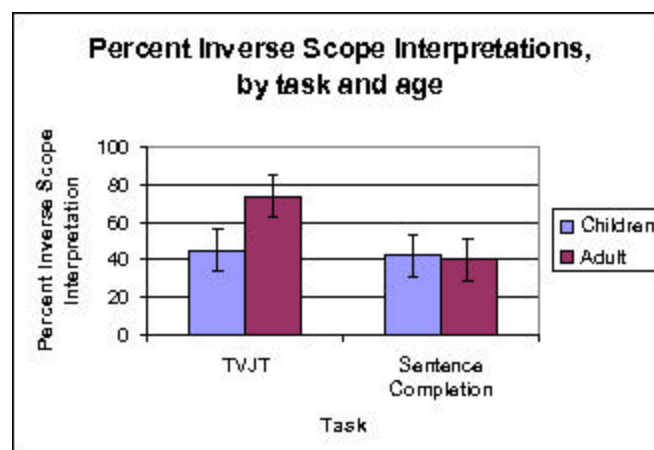


Figure 4.8: TVJT-Sentence Completion: graph of child and adult responses

tween adult rates of obtaining inverse scope interpretations across tasks ($t(19)=4.19$, $p<.001$). Other comparisons were not significantly different.

Discussion

From these results, we conclude that children's interpretations do not vary according to task in the same way that adults' interpretations do. We have seen that adults show higher rates of inverse scope interpretations in a TVJT as compared to a sentence completion task. Children, however, show no such variance. Therefore, it appears that children are lacking the ability to integrate discourse information into selecting an interpretation in an adult-like way.

These results suggest that it is the case that children lack the ability to integrate discourse information into their choice of final interpretation. This suggests that our analysis of the U-shaped curve of development of scope ambiguity resolution is on the right track. That is, children lack the ability to revise their interpretation according to discourse information. However, older children have acquired adult-like parsers, which reduces them to obtaining surface scope interpretations, appearing worse than

their younger counterparts.

4.3 Summary of the Isomorphism Effect

In this chapter, we revisited the isomorphism effect, and found that while we replicated the isomorphism effect with five year olds (as in Musolino et al. (2000)), 4 year olds were able to obtain inverse scope interpretations at quite high rates. This suggested that the advantage for surface scope interpretations (in children) could not be due to difficulty in the parser obtaining the inverse scope interpretation, under the assumption that parsing resources increase with age.

However, we cannot conclude that four year olds are completely adult-like. In Adult TVJT-sentence completion, we found that ‘adult’ behavior can be described not only by obtaining inverse scope interpretations on a TVJT, but also requires modulating one’s interpretation based on task-specific information. We found that children, even if they obtain inverse scope interpretations on a TVJT, do not alter their answers according to task.

This suggests that children lack the ability to use discourse information to resolve ambiguities in an adult-like way. We have proposed an analysis for the U-shaped development that we have observed, suggesting that both children’s parsers and abilities to utilize situational information in interpretation are developing.

At this point we have not determined the source of the advantage for surface scope interpretations in adults. In the next section, we test the parsing hypothesis in adults. We find that a parsing-bias account for adults would be unable to also capture the child development results. Because children develop into adults, it would be desirable to capture the child and adult results under the same system, identifying the exact

factors that differ between the two age groups. Therefore, we suggest that the extralinguistic advantage for surface scope ambiguities is on the right track for adults.

Chapter 5

Testing the Parsing Hypothesis

Our central goal of this dissertation is to identify the processes involved in LF parsing. As we have discussed, there are multiple components involved in the language comprehension process. Therefore, to identify the processes specific to the parsing mechanism, we must we must also identify the relative contribution of extra-linguistic factors to sentence understanding. Only once we have identified the effect of these factors can we determine the affect of parsing decisions on sentence comprehension. In the pursuit of this goal, we have identified two observations that seem related. One, that adults, in online processing appear to have an advantage for surface scope interpretations. Two, there exists a point in children's development when children are limited to surface scope interpretations. With respect to both of these observations, we are attempting to determine whether the advantage for surface scope interpretations is a result of strategies inherent to the parser, or the result of extra-linguistic factors.

Not only is it crucial to determine whether the advantage for surface scope interpretations can be attributed to extra-linguistic factors, as a way of understanding the parser, it is crucial to identify how parsing decisions change over time. That is, chil-

dren eventually develop into adults. Because it is theoretically undesirable to posit that children must acquire a completely new parsing system to become adult-like, it must be the case that there exists a continuity between the interpretations obtained by children, and those obtained by adults. Therefore, if it is the case that children arrive at interpretations in a different way than adults, this must be accounted for by a single description of the language processing mechanism. This view requires that we take data from children and adults as equally informative about the nature of the parser. In this chapter, we propose a model for understanding the contributions of development and extra-linguistic factors to LF ambiguity resolution. Taking the empirical results that we have obtained thus far, we can triangulate the contribution of individual factors to language comprehension. At the conclusion of this chapter, we show that the child and adult results can be accounted for under one model of LF parsing, and that this model requires that the advantage for surface scope interpretations in adults be a result of extra-linguistic factors. Then, given that our model suggests that parsing decisions cannot be responsible for the advantage for surface scope interpretations, we carry out a series of experiments with working memory that confirm this prediction. This model strengthens our description of the source of U-shaped development in the child, but raises the question of which extra-linguistic account can explain the adult advantage for surface scope interpretations.

In the previous chapter, we determined that children's shortcomings cannot be a result of difficulty in the mechanisms underlying inverse scope interpretations, as children are able to obtain inverse scope interpretations at a very young age. We proposed that these results could be accounted for by attributing the U-shaped development to children's developing parsers and abilities to integrate discourse information. We presented an experiment that suggests that children do not integrate discourse information in an adult-like way. In the adult domain, we have determined that both the

parsing and extra-linguistic hypotheses are available, as we have shown that a surface scope advantage can be described by either account. Therefore, in this chapter, we must utilize information from both adult and child processing to develop a unified description of LF parsing. Now, let us attempt to formalize the relative contributions of parser and discourse biases, in a way that captures both the child and adult data.

5.1 Modeling Inputs to Interpretation

In this dissertation, our general goal is to determine the contribution of parsing decisions to the resolution of LF ambiguity. In this section, we build a basic framework for describing the contributions to final interpretation. It is most likely the case that the contributing components identified in this section can be divided into more detailed factors, but we will first match the empirical observations that have been made with a broad division of contributing factors.

Previous research and our experiments have shown that surface scope interpretations are more readily obtained than inverse scope interpretations. An advantage for the surface scope interpretations has been shown both in online processing (Anderson, 2003) and in our own experiments with the IVT and speeded forced choice task. This suggests that, at some level, surface scope interpretations are more economical than inverse scope interpretations. To this point, we have made few claims as to the source of this advantage, but have presented two hypotheses: either a parsing or extra-linguistic advantage. In this section, we suggest that a parsing preference for surface scope interpretations in adults is unable to fit the adult and child data regarding resolution of scopally ambiguous strings.

In addition to determining the source of the surface scope advantage for adults, we have put forth a hypothesis about the child system. In TVJT reinvestigation (section 4.1.1), we found that four year olds appeared to be adult-like in allowing inverse scope interpretations, but that five year olds demonstrated a preference for surface scope interpretations. We suggested that this U-shaped development could be described by a two stage development of the parser and the ability to integrate discourse information. In this section, we show that, using our model of the adult system, we can account for the child results with no modifications to our model. Furthermore, this finding confirms that the dip in adult-like performance observed in children can only be described by an adult system in which the parser has no bias toward surface scope interpretations.

5.1.1 The Basic Model

In this dissertation, we attempt to determine the relative contributions of the different components of sentence comprehension. Specific to the case of scope ambiguity, a string, as in (92), has two possible logical forms, the surface scope interpretation (93a) and the inverse scope interpretation (93b).

(92) Every vulcan isn't as perfect as you think ¹

(93) a. $\forall x[\text{vulcan}(x) \rightarrow \neg \text{perfect}(x)]$

b. $\neg \forall x[\text{vulcan}(x) \rightarrow \text{perfect}(x)]$

From our adult experiments, there are two findings that we must account for. One is that in general conversation, a PF (92) is observed as paired with the logical form in (93b) a higher proportion of the time as compared to (93a) (Musolino and Lidz, 2006). The second finding that we must account for is the advantage for surface scope

¹Adapted from <http://www.startrekfans.net/index.php?showtopic=11359>

interpretations in certain tasks. This advantage is observed when adults are placed in the IVT or speeded forced choice task. These two findings, on the surface, appear to be in conflict. That is, at some level, surface scope interpretations are preferred, but by the conclusion of interpretation, the majority of utterances are interpreted with inverse scope interpretations. Therefore, we must use these two empirical findings as starting points to investigating the contributions of discourse information and parser preferences to final interpretation.

In this section, we will consider two contributors to final interpretation: parsing preferences and discourse information. The term ‘parsing preferences’ refers to any choice made by the parser, independent of discourse information. This may be that the parser prefers to select LFs that are isomorphic to the PF under consideration (Musolino and Lidz, 2005) or to select interpretations with the least complex derivation (Anderson, 2003). If we find that a significant amount of the preference for the surface scope interpretation is carried out by ‘parser preferences’, we may wish to identify the relative contributions of each of these (very different) factors, but we will not for now. The second contributing factor is extra-linguistic information (of which discourse information is a large component). To simplify matters, we will assume that a given situation is deterministic with respect to interpretation.² That is, for a well defined discourse, that discourse yields an interpretation, as in Figure (5.1).

We assume that at start, the grammar makes available multiple possible parses,

²At this point, we have a very poor understanding of the relevant factors that influence interpretation, and so, speculating further would allow us to gain very little. It is possible that a given situation, w_1 , yields surface scope interpretations 60% of the time, and inverse scope interpretations 40% of the time. However, at current, we cannot distinguish this hypothesis from an alternate where w_1 has been poorly defined. Therefore, we must put aside the consideration of non-determinism until we have advanced a more detailed theory of the relation between a discourse context and an utterance.

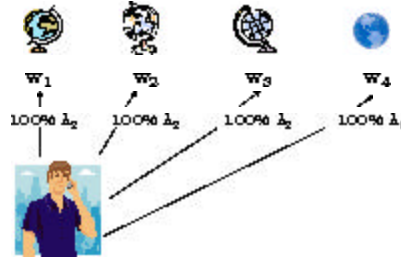


Figure 5.1: The deterministic nature of discourse environments

which will be enumerated λ_1 and λ_2 (there, of course, may be more, but we restrict ourselves to two for simplicity), with equal likelihood.³ We have very strong reason to suspect that parsing preferences and discourse information contribute to language comprehension, and we want to determine how these integrate in order to explain the empirical facts concerning the timecourse of scope ambiguity resolution.

We can ask, given (92), how much more likely is λ_1 than λ_2 ? The degree to which parsing preferences push toward λ_1 over λ_2 will be represented by σ . σ_1 is the probability of λ_1 as the output of the parser, and σ_2 is the probability of λ_2 as the output of the parser. σ is the ratio of σ_1 over σ_2 . This ratio avoids needing to normalize the equation in the course of our computations. This method will make the contribution of each individual component more transparent in the calculations we show. Because the values of σ_1 and σ_2 range from 0 to 1, a value of σ that is greater than one represents a parsing preference for λ_1 , and a value less than one represents a parsing preference for λ_2 . The degree to which extra-linguistic information pushes toward λ_1 over λ_2 will be represented by δ . Just as above, δ_1 is the probability of λ_1 as determined by the integration of discourse information, and δ_2 is the probability of λ_2 as determined by the integration of discourse information. δ is the ratio of δ_1 over δ_2 , again to avoid normalization. Because the values of δ_1 and δ_2 range from 0 to 1.

³This may not be the case, but we will encode the ‘preference for one derivation’ as a parsing preference, and not a product of the grammar.

Again, a value of δ greater than one represents a discourse preference for λ_1 , and a value less than one represents a discourse preference for λ_2 . Additionally, as has been discussed earlier in this work, there exist other types of extra-linguistic information that are not part of the integration of discourse information, such as demands placed by the task. We will discuss these factors when the relevant experimental discussion arises, but we will not build values for these factors into the model. There are likely many different types of extra-linguistic information that fit into this category, and it is not the author's belief that collapsing these factors into one numerical value will provide any illumination on the current investigation.

Let us look at the scope ambiguity case, in which one string has two possible logical forms. We will call the surface scope interpretation of the string λ_1 , and the inverse scope interpretation λ_2 . We can express the overall preference for $\frac{\lambda_1}{\lambda_2}$ as shown in Equation (5.1). A number greater than one represents a push toward λ_1 , while a number less than one represents a push toward λ_2 , and is the product of the contributions of each of the individual components we have outlined above.

$$\frac{p(\lambda_1)}{p(\lambda_2)} = \sigma \times \delta \quad (5.1)$$

Let us run through an example here. Assume that parsing preferences make it two times more likely that the surface scope interpretation will be chosen over the inverse scope interpretation, resulting in $\sigma = 2$. Let us assume that the extra-linguistic information does not contribute any additional information, resulting in $\delta = 1$. Then, the final output of this system ($2 \times 1 = 2$) is that λ_1 is two times as likely to occur as λ_2 .⁴

⁴Of course, here we can wonder what the upper bound on these values are. That is, given a theory where the parser has surface scope as a default interpretation (Lidz and Musolino, 2002; Musolino and Lidz, 2005), σ_1 should be, in theory 0, making σ_1 an infinitely high number. This

5.1.2 Description of the Adult System

We have proposed that there are two components important to final interpretation: parser preferences and the effect of extra-linguistic information. Our first task is to describe the adult system under these parameters. Empirically, we have observed that even though a scopally ambiguous string such as (94) has two possible interpretations, these interpretations are not equally likely across experimental tasks.

(94) Every vulcan doesn't believe in mind melds ⁵

For example, adults interpret sentences like (94) with the inverse scope interpretation 100% of the time in Truth Value Judgment Tasks when the situation makes the interpretation true, but only 30% of the time on the Incremental Verification Task (IVT: TVJT).

One way to determine the relative contribution of parsing preferences is to attempt to look at the output of only that component, and determine if there is an advantage for one interpretation. We can attempt to do this by eliminating the effect of extra-linguistic information. That is, given an empirical value for $\frac{p(\lambda_I)}{p(\lambda_E)}$, and assuming only the value for σ is contributing to this value, we obtain Equation (5.2) (which, is equivalent to Equation (5.1) with $\delta = 1$).

$$\frac{p(\lambda_I)}{p(\lambda_E)} = \sigma \tag{5.2}$$

potential problem reflects on a broader debate about what it means to be a default, whether the parser will always output the surface scope interpretation, or whether it outputs the surface scope interpretation a high percentage of the time. A simple remedy for the current problem is to posit that even if surface scope interpretations are a default, there always exists some slim probability that inverse scope interpretations will be the output of the parser. The exact formulation of parsing defaults must be left for future research by those who support a parsing default account.

⁵Adapted from www.frontierfleet.com/database/xenology.php?id=197

As an attempt to eliminate the contribution of the extra-linguistic component, we developed a task in which the effect of discourse and task information is minimized. TVJTs work under the assumption that an adult will consent to the statement if the statement is true in the situation, creating a response bias in the direction of the true interpretation. This is because this task was developed to test the existence of an interpretation, and cannot be used as any indication of preference. Although intuitively, sentence completion seems as if it makes both interpretations equally available, our sentence completion task involves pictures, which may impose some real-world discourse bias. A task which biases interpretation as little as possible, and requires the least in situational support is a question-answering task, without illustrations. We review this experiment below.

5.1.3 Experiment Seventeen: Non-Biased Reading Experiment

In this experiment, participants are presented with a paragraph (below), that has a scopally ambiguous string in it, and will be referred to as the target sentence (which is presented in **bold** here, but was not differentiated in presentation to the participant). While the paragraph makes both interpretations available, neither is mentioned explicitly, nor is there a bias for either interpretation.⁶ Therefore, this experiment aims at determining the value of $\frac{p(\lambda_1)}{p(\lambda_2)}$, by determining the observed probability of λ_1 , without discourse information calculated as an input.

There was a party at Farmer Jon's farm. A bunny from Hillsdale, a bunny from Stonybrook and a bunny from Camelot came. Farmer Jon offered carrots all around, but they were purple. He also had some cauliflower. Al-

⁶This is in contrast to the sentence completion experiment, where it is explicitly stated that some items are acted on by *none* of the participants, and some are acted on by *not all* of the participants.

though the bunnies were all hungry, each one thought that purple carrots might not taste too good and considered eating the cauliflower instead. But, there was a lot more of the purple carrots and Farmer Jon kept saying how good they were. He really hoped that they would all try them. But in the end, **every bunny didn't eat a purple carrot**. At the end of the day, the bunnies had a cool glass of celery juice to drink.

In this task, we are striving to reduce the discourse biases in interpretation.⁷ Notice, that in this task, the target sentence is embedded in the paragraph. The participant is asked for interpretation after the conclusion of reading the paragraph, by responding to a question. By embedding the sentence in the midst of the paragraph, we attempt to minimize the chances that the participant will recall the exact form of the sentence, and instead recall only the meaning of the sentence. Bransford et al. (1972) found that participants recall gists, and not exact phrasing. This is particularly the case when the target sentence is embedded among other sentences, as remembering the exact phrasing of nine sentences certainly exceeds standard working memory capacity. Therefore, we assume that the target sentence would be encoded only by its logical form (the meaning), and the structure of the sentence would not be stored. If it is the case that the question prompts reanalysis, this reanalysis would not be available. For this reason, we assume that in this task, adults are interpreting the scopally ambiguous string according to their parsing preference.

Following the paragraph, the subject is asked to respond to the question in (95).

(95) Did some bunnies eat a carrot?

⁷Of course, there is much to say here. Because we do not know what factors contribute to adult interpretation, it is difficult to eliminate them. Furthermore, it is difficult to give enough context to make felicitous the use of a scopally ambiguous string, while not biasing either interpretation. We did our best.

If the participant interpreted the target sentence with a surface scope interpretation, they would answer *no*, because it is not true that some bunnies ate a carrot. However, if the participant interprets the sentence with an inverse scope interpretation, they would respond *yes*, because there are some bunnies who ate a carrot.⁸ Therefore, the participants' response to the question reveals how they interpreted the target sentence in the absence of a decisive situation.

Materials and Methodology

Participants were told that they were in a reading comprehension task, and that their task was to read a paragraph and answer a Y/N question about the content of the paragraph. After instructions and practice rounds, subjects were presented with a paragraph on screen. When they had finished reading the paragraph, they pushed the space bar on the keyboard, and a question appeared. The participant indicated their response (either a *yes* or *no* answer) by pressing one of two labeled keys on the keyboard (F and J).

The experiment contained a total of 20 stories: 12 target stories and 8 filler stories. Each target story was presented in one of two variations: either with a negative target sentence (as shown above), or with a positive target sentence. The 12 target stories were followed by either a control or target question. The 8 filler stories were each followed by a filler question. This design resulted in a distribution of the 12 target stories as follows: each participant received three target questions (negative target sentence followed by target question) and nine control questions (positive target sentence or negative target sentence followed by a control question, as described below). It was randomized across participants which question type was associated with which story.

⁸This, of course, relies on the Gricean implicature that *not all* implicates *some*.

The control items were designed to verify that adults were interpreting *some* and *all* as anticipated. An example control question is shown in (96), and appeared after the paragraph presented above.

(96) Did all bunnies eat a carrot?

The expected response to (96) is *no*, because the paragraph explicitly stated that *every bunny didn't eat a carrot* (on either interpretation of the target sentence).

There were two further control conditions, which followed the stories with an affirmative target statement, *every bunny ate a purple carrot*. This paragraph was followed by either of the questions in (95) and (96). In this case, the expected response for (95) was *yes*. However, if a subject made use of the Gricean implicature that *some* means *not all*, then they would answer *no*, and this response would not be scored as incorrect. The expected response for (96) is *yes*, as the paragraph states that *all of the bunnies ate a carrot*.

The filler questions reference parts of the paragraph that were not the target sentence, so that participants could not complete the experiment by simply reading one sentence in the paragraph. An example filler question is shown in (97). Filler stories were presented with either a positive or negative target sentence (just as for targets) to fully mimic the design of the target items.

(97) Was it a cauliflower the bunnies considered eating?

Half of the filler questions had *yes* as the correct answer, and half had *no* as the correct answer.

Participants

20 adult subjects participated in the experiment, with one adult replaced in the design for poor performance on control trials (less than 70% correct). All participants were undergraduates from the University of Maryland. Participants received either payment or course credit for an introductory course for their participation. All participants were native speakers of English.

Results

Across all responses, 48.3% of the trials were answered with the response consistent with the inverse scope interpretation (a ‘no’ response). 9/20 participants obtained surface scope interpretations on all trials. No trials were excluded from analysis.

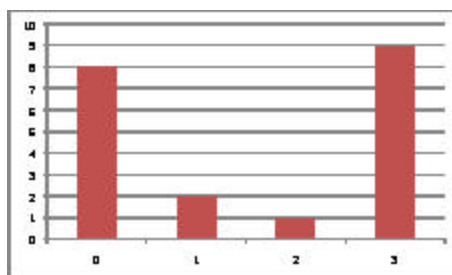


Figure 5.2: Number of Surface Scope Interpretations: Non-Biased Reading Task

A histogram of responses is shown in Figure (5.2). The distribution is largely bimodal, with 17/20 participants obtaining one interpretation in all trials in the experiment.

Performance on control trials was 94% correct. (This calculation excludes *some* control questions following an affirmative target sentence, because both a *yes* and *no* response are appropriate. Including these trials makes performance 96.1% accurate).

Discussion

Adults access the inverse scope interpretation at around chance level, when decisive discourse information is absent. Notice that we observed a bimodal distribution of participants. This is consistent with either one of two possibilities. First, it may be the case that each participant has an around chance probability of picking either interpretation, and in the course of the experiment, the participants maintain the same interpretation. Second, it may be the case that there is a bimodal split in the population, with half of the participants having a parser with a surface scope bias, and the other half with an inverse scope bias. For obvious reasons, this second option is theoretically undesirable unless it is a last resort description of the facts. Furthermore, it would be difficult to formulate the contribution of any type of information on interpretation if for half of the population, the information required revision to surface scope interpretations, and for the other half, triggered revision to the inverse scope interpretation. Therefore, we suggest that the first option is the case, and the bimodal distribution is the result of priming within the experimental context. Therefore, we can infer that the probability of λ_1 is .5. This finding is consistent with our sentence completion results. Therefore, the observation for $\frac{p(\lambda_1)}{p(\lambda_2)}$, assuming no discourse information, is 1.

This forces us to infer that σ is 1. Of course, this does not necessarily mean that there are no parser preferences at all, simply that there is no *overall* parser preference. That is, it may be the case that the parser has a preference for PF and LF pairs that are isomorphic, but that simultaneously, there are language-specific preferences that bias toward inverse scope interpretations, in a proportion that cancels out the overall effect. Because we have not separated out these different types of ‘parser preferences’, we cannot tell here. Our point is that, overall, there appears to be no parser preference for these types of sentences. Therefore, we can furthermore infer

that the preference for inverse scope preferences observed in corpora is a result of extra-linguistic pressures. We would represent this relation in the series of equations presented in Equation (5.3)- (5.6).

$$\frac{p(\lambda_1)}{p(\lambda_e)} = \sigma \times \delta \quad (5.3)$$

$$\frac{.25}{.75} = \sigma \times \delta \quad (5.4)$$

$$.33 = 1 \times \delta \quad (5.5)$$

$$.33 = \delta \quad (5.6)$$

We begin in Equation (5.3) with our baseline equation. In Equation (5.5), we substitute for our observed corpus occurrence (here, 75% observations for inverse scope interpretations. This assumption is spelled out in section 5.2). In Equation (5.6), we solve for the value of δ . Notice that this is a δ that averages over all discourses.

The adult data suggests there is little evidence for an overall parsing preference for surface scope interpretations. If we assume that children begin with architectures quite similar to those of adults, and can pinpoint the sources of change, then we can use our current system to fit the behavior of children. Specifically, recall that in TVJT reinvestigation, we observed that four year olds obtained adult-like interpretations on a TVJT, but that five year olds reverted to surface scope interpretations. We suggested that these data could be accounted for by children beginning with non-adult-like parsers, and furthermore, acquiring the ability to integrate dis-

course information. Now, we will show that this suggestion indeed functions under our current system.

5.1.4 Description of the Child's System

To describe the U-shaped pattern of development in children's ability to resolve scope ambiguities, we proposed two factors in children's development. We suggested that both the parser and children's abilities to integrate discourse information were developing. We suggested that four year olds had non-adult-like parsers, in that they contained a bias for inverse scope interpretations. At the next stage, we suggested that children obtained adult-like parsers, but lacked the ability to alter their interpretation according to discourse. Adults have adult parsers (which we have determined have no parsing bias for either interpretation), and the ability to integrate discourse information. In this section, we will show that, using the system we have already outlined, these assumptions about U-shaped development indeed describe the data.

In our model, we have two contributors to interpretation, parsing preferences and extra-linguistic information. For simplification, we will assume that the largest contributing component of extra-linguistic information is discourse information, and we will speak of children's abilities to integrate discourse information. However, in fact, children are most likely acquiring other extra-linguistic skills, as well. Our first claim about children is that they do not use discourse information to disambiguate scopally ambiguous strings in adult-like ways (Trueswell et al., 1999; Musolino and Lidz, 2006). To model this, we will remove the contributions of the discourse information (and extra-linguistic information) from the equation.⁹ To model children, we will

⁹It is an open question as to whether this component is underdeveloped, or simply if the weights this component contributes to the selection constraint are not set. In any case, this will play the same explanatory role for us.

make the assumption that children lack the contribution of δ .

From TVJT reinvestigation, we found that 4 year old children behave adult-like in a Truth-Value Judgment Task, but 5 year olds do not. Children are presented with a story, and a target sentence in (98).

(98) Every mouse didn't eat a piece of cheese

In this task, adults assent to the statement (consistent with the inverse scope interpretation) 76% of the time. The four year olds, 4;5-5;2 (mean 4;9), give a response consistent with the inverse scope interpretation 78% of the time, not significantly different from adults. 5 year olds, 5;3-5;7 (mean 5;4) however, give inverse scope interpretations 53% of the time,¹⁰ significantly less often than the four year old children ($t(28)=2.05$, $p < .05$).

Therefore, four year olds are adult-like in their performance on this task, while 5 year olds are not. We suggested that the difference between four and five year old children was a difference in parsing preferences. Let us make this idea more concrete.

We have claimed that four year olds have non-adult-like parsers. That is, four year olds obtain their inverse scope interpretations via a parser that is biased towards inverse scope interpretations. We will assume the child begins with a grammar that allows for two interpretations for a scopally ambiguous string as in (98), as it has been shown that children's deficits are not grammatical in nature (Gualmini, 2004; Musolino and Lidz, 2006). We will additionally assume that the child attempts to

¹⁰We can say that at year 5, only half of the children have entered this phase, and drop to 100% isomorphism (single generation). Alternatively, we can say that all kids hit this drop, and because there is dual generation, 50% of the children get inverse scope (on the first trial), and 50% get surface scope.

match their interpretation (that is, the selection of an LF, given a PF) to what they observe in their input by adjusting σ . Assuming that the child receives standard adult usage in their input, this yields Equation (5.7).¹¹

$$.33 = \sigma \tag{5.7}$$

We have a child who, with limited resources, attempts to match their language production system such that, when encountering a PF, it aligns with the proper LF, at the rate in which this occurs in the environment. Of course, this assumes that a child is able to perceive which logical form was intended by the adults around them, which is most likely not a benign assumption. However, if we are to assume there exists a contribution of discourse information to an interpretation, then this must be learned in some way. Therefore, while it is clear that the observation problem is a difficult one, we put it aside for now. This non-adult-like parsing bias (without any input of discourse information), then, can describe the young children's preference for inverse scope interpretations.

We have claimed that five year olds obtain the adult-like parser (that is, the adult parsing bias, or lack thereof), but still have no ability to integrate discourse information into their selection of an interpretation. As we have determined that adults have no parsing bias, it then follows that five year olds, in obtaining an adult-like parser, would have no parsing bias. This yields the equation in Equation (5.8).

$$.5 = \sigma \tag{5.8}$$

This results in 5 year olds obtaining surface scope interpretations 50% of the time,

¹¹We assume that the child observes 75% inverse scope interpretations. This corresponds to .33 (1/4 observances are surface scope, resulting in a 1 to 3 ratio of surface scope interpretations).

the observed result. Therefore, we suggest that the five year olds can be described as suggested by our U-shaped analysis, by having adult-like parsers. However, children of this age group are lacking the ability to use discourse information as a means of influencing their choice of interpretation. Of course, it seems clear that the notion of ‘discourse information’ must be broken down into further components, as it is certainly the case that children of this age can participate in conversational acts, and do not entirely ignore information about their interlocutors. The correct analysis is most likely that there are some components of discourse information that children are sensitive to, and some components that children of this age are not sensitive to, and that there is a correlation between the factors they are not sensitive to, and the factors that guide selection of an interpretation in adults. For now, we will oversimplify, and call the range of components ‘discourse information’. This is not meant to imply that there is nothing more nuanced to this notion.

To become adults, children must acquire the ability to integrate discourse information into their selection of an interpretation. At this point, we have no account of how this works. It may be the case that children had the ability to observe discourse all along, but simply could not use the information to inform their choice of an interpretation. Here, we will simply outline an idea for how a child may notice that they need to utilize discourse information to match adult usage.

When the child only has σ , they are computing the likelihood of LFs, given a PF, without the ability to use this information contributed by the discourse context. That is, the child takes the observed frequency of PF/LF pairs, and distributes it across worlds, as shown in Figure (5.3).

The result is that the child, essentially, has a probabilistic mechanism for determining scope assignment independent of discourse.

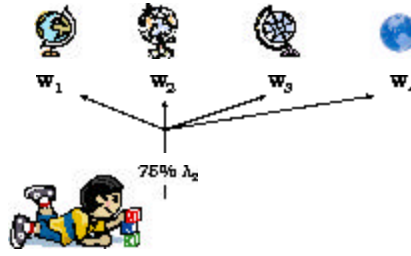


Figure 5.3: The child's view of the discourse context

The child will obtain the same interpretation, regardless of the discourse environment around them. Because this is not how the adult system works, we must describe how the child would become adult-like. This is a two-step process. First, the child has to notice that there exist different discourses, but is still unable to attribute linguistic relevance to them. Recall, there are some worlds for which adults obtain always surface scope interpretations, and some for which they obtain only inverse scope interpretations. At some point, the child will realize that in a discourse, w_1 , they will predict 25% surface scope interpretations (as they have distributed the percentage over situations), but adults in that discourse always obtain inverse scope interpretations, as in Figure (5.4).

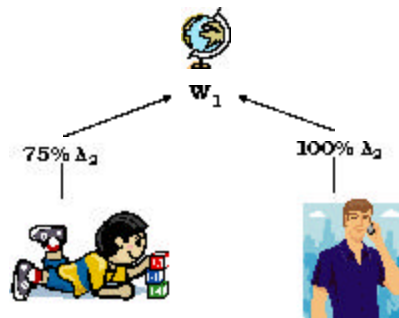


Figure 5.4: Child and Adult Comparison of Interpretation

It is then that the child will realize that they must alter their interpretation according to discourse, and therefore, utilize δ . That is, the child will be able to use discourse information to alter their interpretation.

5.1.5 Conclusions

We began investigating the advantage for surface scope interpretations in both children and adults. From our experiments, we determined that the advantage for surface scope interpretations observed in adults could not be explained by a simple parser bias for surface scope interpretations. We created an experiment that minimized the contribution of discourse information to the experimental situation. In this experiment, we found that adults were at chance with respect to interpretation, in line with our sentence completion results. This suggested that overall, there are no parsing preferences that contribute to final interpretation. This suggests that the generation of alternatives (at least, LF alternatives) does not give any privileged status to either interpretation. This suggests that the advantage for surface scope interpretations found in the IVT and speeded sentence completion tasks is an extra-linguistic effect.

Furthermore, we reviewed our initial analysis of the U-shaped pattern of development of ambiguity resolution in children. We showed that an analysis in which children began with biased parsers, and later acquired adult-like parsers can describe the empirical evidence. Importantly, we were able to model the experimental results using the same system that we posit for adults. This suggests that our analysis is on the right track. Determining how it is that children acquire the ability to integrate discourse information into final interpretation is a question for future research.

Importantly, this account has allowed us to formulate our intuition that the adult and child parsers must be related, with changes in only a small number of factors. This is a desirable finding, and opens the possibility that there are aspects of sentence processing we can only pinpoint by combining empirical evidence from both children and adults.

5.2 Assumptions about Corpora and Inverse Scope Interpretations

In the previous section, we assumed that the observed rate of inverse scope interpretations for scopally ambiguous strings is 75%. Ideally, we would search corpora, code for interpretation, and arrive at an exact number. However, for reasons we describe below, determining an intended interpretation is not so straight-forward. Therefore, we have chosen the number 75% to capture the previous reports that surface scope interpretations are preferred, but that surface scope interpretations do occur (Musolino and Lidz, 2006). Our claims would not be changed if this number actually turns out to be 65% or 90%, as long as it is a majority of the time.

As Musolino and Lidz point out, there are certainly a large number of cases in which one can be sure inverse scope is intended, just from text alone. One such example is in (99).

- (99) Of course, everybody didn't get conned. Sherrod Brown didn't. Dennis Kucinich didn't.

mcdac.blogspot.com/2007/05/voinovich-on-iraq-everybody-got-conned.html

Because the scopally ambiguous string is followed by a list of people who didn't get conned, we infer that (99) means *not everyone got conned* and not the surface scope interpretation *no one got conned*.

There are also cases where the surface scope interpretation is clearly intended, as in (100) and (101).

- (100) I decided to stay down here because everybody didn't think I should commute.
Whenever I told people I might not consider moving down here, everybody

had the same shocked expressions.

http://www.jackiechan.com/message_view?cid=642

(101) You just have to make sure everyone on your committee isn't out of town

Ellen Lau, on picking a defense date

In these cases, either the following text or real world knowledge makes the surface scope interpretation clear. In (100), the sentence following the scopally ambiguous string confirms that everyone shared the opinion that Jackie Chan should not commute. Real world knowledge (that one holds defense dates with all of one's committee present) forces us to interpret (101) with the surface scope interpretation *none of your committee is out of town*. An informal survey by the author found numerous examples of such surface scope interpretations, suggesting the 'one of dozens' found by Musolino and Lidz may be a low estimate.

Despite the presence of these clear cases, many scopally ambiguous strings are difficult to interpret. This seems contrary to our intuition, since a listener, when encountering such a string, has no trouble arriving at a meaning. However, in many cases, both the surface scope and inverse scope interpretations are consistent with the situation. Let's look at an example from Musolino and Lidz that is claimed to have an inverse scope interpretation, shown in (102).

(102) Everybody doesn't have to be the same

Musolino and Lidz (2006)

The two interpretations for (102) are *not everyone has to be the same* or *no one has to be the same*. While it is certainly the case that the inverse scope interpretation is true, it may also be the case that the surface scope interpretation is true. There is no information present that decisively selects an interpretation. In fact, in conversation,

the actual interpretation may not serve an important discourse function. That is, it is possible that different listeners interpret (102) with different interpretations, and this seldom has an effect on the following discourse.

Another reason that interpretation can be difficult to pinpoint is that the domain of quantification is determined by the intentions of the speaker. In example (100), it is probably not true that *everyone in the United States thought Jackie Chan shouldn't commute*. Likely, it meant that *everyone that Jackie Chan talked to thought that he shouldn't commute*. If someone hearing (100) thought the 'everyone' meant *everyone in the United States*, then the sentence would be true only on the inverse scope interpretation. However, in this case, it was intended to be true under the surface scope interpretation. Therefore, having the incorrect domain of quantification could also give rise to inappropriate scope interpretations.

Due to the vast number of sentences like (102) (see (103-104) for more examples¹²), we do not believe it is possible to obtain a fair estimate from a corpus for inverse scope interpretations.

(103) Everybody didn't know what was happening.

www.usagi.org/doi/digiko/nyo/47.html

(104) What should we do when everything doesn't go our way...?

drjazzi.blogspot.com/2007/07/when-everything-doesnt-go-your-way.html

(105) Everyone doesn't like everyone

preggiepoos.blogspot.com/2007/10/everyone-doesnt-like-everyone.html

Therefore, in this work, we use the estimate of 75% inverse scope interpretations, because we believe it is, in principle, impossible to obtain a fair estimate of inverse

¹²All web citations were found using Google.

scope interpretations from a corpus. This raises the question of how children learn scope preferences in the adult language, and what role these ambiguous sentences play. We leave this issue for future research.

This issue aside, we would like further experimental confirmation that the advantage for surface scope interpretations in adults are not the result of a parsing preference. In the next section, we turn to using working memory as a tool to investigate the factors that contribute to final interpretation.

5.3 Working Memory

S-structure parsing involves the construction of a representation from a string of words. In order to build a constituent structure, a listener must hold the words that constitute that structure, and then compose those words into an s-structure representation. This process must be done within the cognitive capacities of the listener, because all of language is processed through a human mind. The requirements on the construction of a LF are no different. That is, even assuming an s-structure representation, the listener must derive an LF tree from this representation, implementing the manipulations necessitated. In the domain of s-structure parsing, it has been shown that limitations on a listener's working memory impair their ability to build s-structure representations. That is, it has been shown that adults with lower working memory capacities have difficulty maintaining multiple interpretations of an s-structurally ambiguous sentence (Macdonald et al., 1992). This suggests that because adults with lower working memory capacities have less working space available for the building of s-structure representations, they are restricted to building only the simpler available of multiple interpretations. It is also possible to replicate the effects of limited working memory capacity by exposing adults to a concurrent task, which

temporarily limits their working memory capacity.

If it is the case that there exists inherent difficulty in the construction of inverse scope interpretations, then we expect the same results to arise in this domain. That is, if it is the case that the building of the logical form associated with the inverse scope interpretation of a scopally ambiguous string is more difficult than building the logical form associated with the surface scope interpretation, then we expect working memory limitations to impact an adult's ability to obtain inverse scope interpretations. More specifically, in a task where an adult is given a concurrent load, which presumably reduces their working memory capacity, we expect the adult to obtain high rates of surface scope interpretations.

However, if it is the case that inverse scope interpretations pose no additional processing cost with respect to the parser, as we have suggested, then we expect there to be no effect of working memory manipulations. Therefore, we want to determine if the 'difficulty' associated with obtaining inverse scope derives from factors internal to the parser or from extra-linguistic factors. Such extra-linguistic factors may include discourse information or verification of the interpretation with respect to the world. In this section, we utilize the non-biased reading task, a task that allows for interpretation without a rich discourse, allowing us to empirically test between the parsing and extra-linguistic hypotheses. If it is the case that working memory resources are taxed in the process of obtaining inverse scope, then we expect adults to obtain surface scope interpretations at higher rates under a working memory load, due solely to the decreased parsing resources available. However, if it is the case that the difficulty can be described by the consumption of working memory resources during the integration of discourse information, then we expect that working memory load, in this task, should reveal no interpretive effect.

Let us first outline our basic assumptions about working memory. We will be working within Baddeley's model of working memory (Baddeley and Hitch, 1974; Baddeley, 1986). We have chosen this framework because it is the one in which the most research on language comprehension has been conducted, not because we require any specific components of the architecture. What is crucial to us is that the model assumes limited processing resources, and that these are the same resources required in maintaining attentional focus.¹³

Baddeley's model has a central executive component which has a limited amount of processing space. There are two slave systems, the phonological loop and the visuo-spatial sketchpad, as sketched in Figure (5.5).

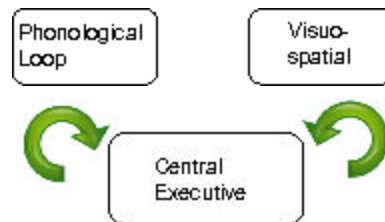


Figure 5.5: Sketch of Baddeley's Working Memory Model

The phonological loop is essential for speech perception, and the central executive has been implicated in higher level language comprehension (Baddeley et al., 1984). The phonological loop is required for maintaining the lexical items in the speech stream, until the hierarchy of an s-structure representation can be built. Therefore, the phonological loop plays an important role in language comprehension. However, the capacity of the central executive system also plays an important role in language comprehension. It has been shown that limitations of central executive will lead

¹³This could also be captured by Cowan's model, in which there is a focus of attention and activated long term memory, each component which has processing limits (Cowan, 1999). The logic would still follow.

to poorer performance on other tasks that utilize the same resources, like language processing. With language comprehension, previous research has shown that low span adults have difficulty in the processing of syntactically complex sentences. One example is a study done by King and Just (1991), which found that participants with low working memory spans had more difficulty with syntactically complex sentences than those with medium or high spans. King and Just compared participant's reading times of strings containing subject (106) versus object (107) relative clauses.

(106) The reporter that attacked the senator admitted the error

(107) The reporter that the senator attacked admitted the error

Strings containing object relative clauses are more complex because *reporter* is both the subject of the main clause and the object of the embedded clause. King and Just measured reading times of these strings with low, medium and high span participants. They found that low span participants were slowed down more in the object relative clause case than the subject relative clause case, particularly at the second verb, which is posited to be a point of increased working memory difficulty. Specific to the case of ambiguity resolution, Macdonald et al. (1992) showed that low span adults have difficulty maintaining multiple interpretations, as compared to high spans. This suggests that a limited working memory capacity impairs one's ability to fully process syntactically ambiguous strings.

Taxing the central executive can occur in a variety of ways: a) the participant may have limited resources to begin with (low-span participants), b) high syntactic complexity of the material being processed or c) concurrent load (Carpenter et al., 1994). For example, Waters et al. (1987) found that articulatory suppression affected adults' ability to process syntactically complex sentences.

For the investigation of LF ambiguity resolution, this logic easily translates. It seems that ambiguity resolution requires resources, and we want to know if scope ambiguity patterns in the same way. The parsing hypothesis claims that obtaining inverse scope interpretations is difficult, and requires central executive resources (put differently, scopally ambiguous strings are just one specific type of a syntactically complex sentence). If this is the case, then we expect that accessing of inverse scope interpretations will be more difficult in a situation where there is a concurrent load.

We will have participants participate in the non-biased reading task, while performing concurrent load tasks. Creating a concurrent load task, such as having participants listen to irrelevant noise, eliminates the need to do independent memory span tasks on each participant. Assuming that a concurrent load task limits the amount of working memory resources available for sentence comprehension, we can determine whether scope interpretations are affected by limited working memory. If we observe that surface scope interpretations increase under load, then we can infer that obtaining inverse scope is more difficult, by virtue of the nature of the representation. However, if we find no evidence that load affects interpretation, it would appear to be the case that inverse scope interpretations are not more complex than surface scope interpretations. If this is the case, then we may conclude that the surface scope advantage is due to extra-linguistic factors, and not as a result of parsing difficulty associated with the syntactic complexity of the representations.

5.3.1 Experiment Eighteen: Artificial Language Word Span Experiment

It has been shown that irrelevant speech disrupts the accuracy of recall of a word list, suggesting that the speech is processed, and overwrites information that should have been stored (and rehearsed) in the phonological loop (Colle and Welsh, 1976). We

have presented some research that suggests that disrupting the information stored in the phonological loop also disrupts higher-level language processing. We want to use this finding to determine whether disrupting speech affects the process of scopal ambiguity resolution. First, however, we must determine that we have identified a sample of irrelevant speech that will disrupt storage in the phonological loop, to replicate previous results.

In this experiment, we attempted to tax the phonological loop by playing participants artificial language stimuli as we tested word span. We used artificial language which approximated English, under the assumption that this would be processed, and use up central executive resources (because we are working under a limited processing capacity model). If it is the case that this stimuli disrupts the participant's ability to recall a word list, then we can use this same stimuli as a concurrent load task in testing the resolution of scope ambiguity.

Materials and Methodology

Artificial language stimuli were artificial one-syllable words, presented in random order by a Psyscope script.¹⁴ These non-words followed the rules of English phonology, and were plausible non-words.

The word span task tested the length of a list the participant could recall. Two sets of lists were created for the experiment. Each set contained lists of words which ranged in length from one to nine words, in increasing order. Words were two syllable words chosen at random.¹⁵ The lists contained no overlapping words, and no list contained words which were clearly semantically or phonologically related.

¹⁴The stimuli were created independently for an artificial language experiment by Eri Takahashi.

¹⁵All words selected were found on cnn.com, and therefore, were assumed to be in common usage.

Participants were told that their task was to recall as many words on the list shown to them, in order. Participants were given 1 second per word on the list to memorize the words, followed by 1 second where the list would disappear. During this time, the word *wait* appeared on the screen, to indicate that the participant should not begin recall. After this time, the participants were prompted to recall the list (in a written format) on a piece of paper. Participants had an unlimited amount of time to recall. When the participant had finished recalling the items, the participant would progress to the next list. This continued until the participant could no longer recall all of the words on the list.

Each participant participated in two conditions of the experiment. In one condition, the task was presented exactly as described above (the no-sound condition). In the second condition, the participant performed the task above while a loop of artificial language played concurrently (the sound condition). List and condition combinations were counterbalanced. Two lists of words were created, as described above: List A and List B. Participants were divided into four groups: Group one, two, three and four. Group one and two were given List A first, and Groups three and four were given List B first. Groups one and three began in the sound condition, while Groups two and four began in the no-sound condition.

The entire session lasted for 2 sessions of approximately 5 minutes each. The two sessions were separated by approximately 30 minutes by an unrelated filler task.

Predictions

If it is the case that artificial language disrupts storage in the phonological loop, and therefore, taxes working memory, then we expect participants in the artificial language condition to recall fewer words than in the no-artificial language condition.

If it is the case that we have not found an adequate task to tax working memory, then we predict there to be no effect between conditions.

Participants

20 adults participated in this experiment. No participants were excluded. Participants were graduate or undergraduate students from the University of Maryland, and received either course credit or monetary compensation. All participants were native speakers of English over 18 years of age.

Results

In the no-sound condition, participants recalled an average of 4.65 words. In the sound condition, participants recalled an average of 3.65 words. The two conditions are significantly different ($t(19)=-3.98, p<.001$).

Discussion

It appears that the artificial language, which, by hypothesis, would require phonological loop processing, had an effect on word span. Therefore, we can be confident that we have identified a set of stimuli that impacts phonological loop storage. This finding conflicts with a study performed by Martin et al. (1988), who found that non-speech had the same effect as white noise on reading comprehension (a null effect), while speech impacted comprehension. It is possible that the artificial language presented in this experiment served as a more accurate representation of genuine speech, and therefore caused a similar disruptive effect as speech. Given that we have determined that this artificial language disrupts phonological loop storage, we will use this stimuli in the next task to test the effect of irrelevant speech on the resolution of scope ambiguity.

5.3.2 Experiment Nineteen: Artificial Language and Scope Experiment

This experiment attempts to tax the phonological loop, as a means of limiting the amount of processing space in working memory. This works in two ways: by eroding the information in the phonological loop, making it unavailable for subsequent retrieval, and by increasing the amount of processing required in the phonological loop, draining resources from the central executive system (because we are working in a limited capacity model). With respect to the first point, Baddeley et al. (1987) claim that the phonological loop is used as a backup for storing sentences that require further analysis. While simple strings are comprehended in real time, some complex sentences (such as semantically reversible strings and those containing center embedded clauses, as in (108)) require further semantic analysis, requiring a stored representation of the sentence to refer back to (Saffran and Martin, 1975).

(108) The Klingon that was kicked by the Vulcan had a black shirt

The parsing hypothesis assumes that the surface scope interpretation is a default, and that accessing inverse scope requires revision. Therefore, the parsing hypothesis requires that revision occur. If revision occurs the same way for scopally ambiguous strings as for syntactically complex sentences, then we would expect that eroding the phonological loop may prevent this revision. As we have seen, one way to tax the phonological loop is with irrelevant speech. Colle and Welsh (1976) found that irrelevant speech disrupts the recall accuracy of a word-list, suggesting that the speech is processed, and overwrites information that should have been stored (and rehearsed) in the phonological loop. With respect to the second point, decreasing the amount of processing space in the central executive eliminates participant's ability to maintain multiple interpretations of an ambiguous sentence, as discussed above. Therefore, the presence of a concurrent task that may impact the availability of processing space,

under the parsing hypothesis, should also limit a participant's ability to maintain the more difficult, inverse scope interpretation.

In this experiment, we give adults the non-biased reading task. We attempted to tax the phonological loop by playing participants artificial language stimuli as they performed the task, which subjects reported as mildly disruptive, but tolerable. We used artificial language which approximated English, so that the language would be continually processed by the phonological loop.

Materials and Methodology

Artificial language stimuli were artificial one-syllable words, presented identically to those used in the previous experiment. Participants were introduced to the non-biased reading task the same as previously described. After completion of describing the task, they were told that they would be listening to some noise while trying to perform the task.

Similar to the baseline experiment, targets were scored based on Y/N responses. The entire session lasted approximately 20 minutes.

Predictions

We have been considering multiple potential sources of the advantage for surface scope interpretations. One hypothesis is that the parser gives preference to surface scope interpretations, and that inverse scope interpretations can only be accessed at an increased processing cost. The second hypothesis is that surface scope interpretations gain an advantage, not in the linguistic domain, but due to some extra-linguistic source.

Previous research in the domain of working memory and language comprehension lend motivation that our assumptions concerning the predictions of the parsing hypothesis are correct. It has been shown that sentences that are syntactically more complex are more difficult to comprehend (King and Just, 1991; Macdonald et al., 1992). Therefore, if inverse scope interpretations are more complex, then it seems likely this interpretation requires an increased amount of resources. The parsing hypothesis predicts that the accessing of inverse scope interpretations is difficult, and that increasing working memory load while reading scopally ambiguous strings should increase surface scope interpretations. If it is the case that the surface scope interpretation is a default interpretation, and the inverse scope interpretation requires extra effort to access, then we would expect an increase in surface scope interpretations in this task over the baseline condition.

However, if the extra-linguistic hypothesis is correct, then we expect to see no increase in surface scope interpretations. That is, if it is the case that there is no inherent increased difficulty associated with obtaining inverse scope interpretations, then we expect working memory manipulations to have no effect.

Participants

20 adults participated in this experiment. No participants were excluded. Participants were undergraduate students from the University of Maryland, and received either course credit or monetary compensation. All participants were native speakers of English over 18 years of age.

Results

Across all target trials, participants obtained surface scope interpretations 58.3% of the time. This percentage is not different from the baseline condition ($t(38)=-$

0.44, $p=.65$). 10/20 participants obtained surface scope interpretations on all trials. Performance on control items was 93% correct.

Discussion

It appears that the artificial language, which, by hypothesis, would require phonological loop processing, had no impact on the rate of surface scope interpretations obtained. This is consistent with the hypothesis that there is no preferred interpretation, and accessing inverse scope interpretations requires no extra work. Recall that impairing phonological loop storage requires not only utilizing some central executive resources in a limited resource system, but also eliminated the possibility for subsequent processing of the verbal material. This, alone, casts doubt on the view that accessing inverse scope interpretations requires revision (and that surface scope interpretations are a parsing default), because it is unlikely that revision would be able to occur in this case. However, to be confident in our results, we turn to an experiment which uses a concurrent load task that directly limits working memory capacity.

5.3.3 Experiment Twenty: One Word Load Experiment

Word load, which requires processing space, may be a more direct limitation to central executive working memory capacity. Word load, that is, remembering a word for later recall while performing a task, is a common tool for taxing working memory capacity. King and Just (1991) required participants to maintain either one, two or three words in memory while reading a range of syntactically complex sentences. They found an interaction between size of memory load and syntactic complexity in comprehension. This suggests that working memory load impacts sentence comprehension.

In this study, participants are required to read an entire paragraph, aloud. Previous experiments require adults to recall a word over a single sentence, and have found effects with three word load. However, because we require participants to read an entire paragraph, we lowered the word load accordingly to one word. Additionally, it has been shown that articulatory suppression prevents rehearsing (Estes, 1973). Therefore, the reading of the paragraph aloud has two functions. First, because the paragraph is read aloud, the participant cannot rehearse the word they are supposed to recall. Second, the reading aloud of the paragraph also acts as articulatory suppression of the previous sentences. Once the scopally ambiguous string is encountered, the participant will continue reading the next sentence in the paragraph. Therefore, the next sentence restricts the participant from additionally processing the scopally ambiguous string.

Materials and Methodology

In this task, the non-biased reading experiment was presented on the computer and described to the participant exactly as before, except that they were asked to read the paragraphs aloud (this is to prevent rehearsal, for the reasons described above). Participants were recorded to verify that the paragraph were read aloud. Participants were told that they would be concurrently conducting a memory task. They were given a sheet of paper, and told their task was to remember the last word of the previous paragraph. After answering the question about the current paragraph, they were to recall this word on the sheet of paper.

Trials are scored according to the previous criteria. The entire experiment lasted approximately 25 minutes.

Predictions

The predictions for this experiment remain the same as for the previous experiment. If it is the case that there exists a parsing advantage for surface scope interpretations, then we should expect an increase in the number of surface scope interpretations obtained in this experiment. This is for two reasons. One, we have added a word load, which draws on processing resources available. Two, the articulatory suppression caused by the reading aloud prevents further post-processing of the scopally ambiguous string, assuming that later analysis is required by accessing the phonological loop store (Baddeley et al., 1987).

However, if it is the case that there is no additional difficulty associated with accessing inverse scope interpretations, then we predict that the current task will have little effect on interpretation. That is, if adults process the inverse scope interpretation with no increased difficulty over the surface scope interpretation, then word load and articulatory suppression should have no effect on this process.

Participants

20 adults participated in this experiment. No participants were excluded. Participants were undergraduate students from the University of Maryland, and received either course credit or monetary compensation. All participants were native speakers of English over 18 years of age.

Results

Across all trials, participants obtained the inverse scope interpretation 61.7% of the time. This result is not significantly different from baseline ($t(39) = .845$, $p = .40$).

5/20 participants obtained surface scope interpretations on all trials.¹⁶ Performance on control trials was 90.8% accurate.

Discussion

In this experiment, we added both a concurrent load and an articulatory suppression component, in an attempt to tax on working memory capacity. However, we found no effect on interpretation. Therefore, we conclude that there appears to be no increased difficulty associated with accessing inverse scope interpretations.

5.3.4 Conclusions the Parsing Hypothesis

At present, it appears that there is no effect of working memory load on interpretation of scopally ambiguous strings. Recall that a parsing hypothesis posits an increased cost for accessing inverse scope interpretations. However, even under concurrent load measures, we fail to find any increased cost for obtaining inverse scope interpretations. Therefore, we suggest that these results lend support to the extra-linguistic hypothesis. These findings are compatible with the results from our model, which suggests that there is no initial benefit for surface scope interpretations. That is, it seems unlikely that accessing inverse scope interpretations require revision, as we would have expected such a process to be hindered by constraints on working memory resources.

¹⁶It should be observed that 9/10 participants obtained solely surface scope interpretations in the baseline experiment, which appears to differ from the current results. It is important to note, however, that even if one takes this distribution to be informative, then it suggests that participants are more likely to access inverse scope interpretations in this task over baseline. This, if true, presents further evidence against the parsing hypothesis.

We conclude that there is no parsing advantage for surface scope interpretations, suggesting that the source of the advantage for surface scope is extra-linguistic. According to our model, it cannot be the case that discourse factors are responsible for the advantage for surface scope interpretations, because in conversation, it is the case that adults adhere to inverse scope interpretations. However, recall that there are other extra-linguistic factors that we did not take into account in our model. In the next chapter, we suggest that the verification procedures associated with the verification of interpretations, gives rise to the advantage for surface scope interpretations. Although this will be outlined in greater detail, let us preview how this account would work.

We have determined that the parser makes available multiple alternative interpretations of scopally ambiguous strings, with no advantage given to either interpretation. Therefore, to obtain a final interpretation, extra-linguistic factors must exert pressures to select an interpretation. We suggest that the verification procedure associated with the surface scope interpretation is easier than that associated with the inverse scope interpretation. Therefore, given a case where both interpretations are made equally available by the parser, the surface scope interpretation will be selected, whenever these verification procedures are required. Under this account, we must explain why adults obtain inverse scope interpretations in the non-biased reading task about half of the time. That is, if surface scope interpretations gain an advantage due to verification procedures, then why do they not in this task? Here is where a difference between the non-biased reading experiments and the experiments where we find an advantage for surface scope interpretations must be pointed out. Notice that in the non-biased reading experiments, adults read and interpret a paragraph based on mental representations of the world, and are not required to verify the interpretations with respect to the real world. That is, in this experiment, where no verification

is required (at least, in the real-world sense of verification) there is no advantage for either interpretation, because the grammar makes both interpretations available, and the parser selects between them. However, in real-world verification procedures, there is an advantage for the verification procedure for surface scope, over inverse scope interpretations. Therefore, in any task where there are real-life pictures, it will appear that there is an initial preference for surface scope interpretations. These, generally, can be overcome with the integration of discourse information, except in tasks where the integration of discourse information is restricted. Therefore, it seems initially plausible that verification procedures can account for the observed advantage for surface scope interpretations.

In the next chapter, we will investigate verification procedures as the source of the preference for surface scope interpretations.

Chapter 6

Verification

To this point, we find overwhelming empirical evidence for an advantage for surface scope interpretations, but we find no linguistic motivation for such a preference, and fail to find increased rates of surface scope interpretations under a working memory load, which presumably, is one of the best measures we have for ‘general cognitive difficulty’. These results suggest that extra-linguistic factors are the source of the advantage for surface scope interpretations. We have sketched a proposal for how a difference in verification procedures for the two interpretations would account for such an advantage. In this chapter, we illustrate in greater detail how verification strategies can be used to describe the phenomenon we have observed. We claim that the preferences we have seen fall out from this asymmetry in verification preferences, not from facts about the grammar or the parser. At the conclusion of this chapter, we will demonstrate how this account extends to a variety of findings. We show that not only does this account predict an advantage for surface scope interpretations for sentences containing a universally quantified subject and negation, but it also accounts for the findings concerning sentences containing number and negation, and the interaction between two quantifiers, as demonstrated in Anderson (2003). This chapter presents an account attributing linguistic findings concerning semantic ambiguity resolution

to independently required psychological phenomenon.

Identifying the contribution of verification procedures to interpretation is a crucial component to studying language comprehension. After all, the study of most types of constructions require verification in the experimental setting. Therefore, the finding that these issues are relevant to the selection of an interpretation bears on a wide variety of linguistic investigations.

6.1 On Verification Procedures

Besides the fact that verification is one extra-linguistic phenomenon which unites the variety of experiments we have reviewed, there is reasonable evidence that not all verification procedures are created equal. Research from the psychology literature demonstrates that while adults are adept at performing some classes of verification procedures, they fail with others (Wason, 1965). This suggests that some verification procedures are more difficult to employ than others.

Adults find a sentence like (109) relatively easy to verify.

(109) If something is a snail, then it has antennae

That is, in a world where there are four snails, adults easily carry out the procedure of checking each snail, and seeing if it has antennae. If it is the case that each snail does in fact have antennae, then the statement is true. If there is a snail that fails to have antennae, then the sentence is false. Notice that the sentence in (109) is logically equivalent to the sentence in (110), and can both be written as (111).

(110) Every snail has antennae

(111) $\forall x[\text{snail}(x) \rightarrow \text{antennae}(x)]$

The logical form in (111) is verified by the procedure described above.¹ We are assuming that the verification procedure that is carried out is very tightly related to the statement of logical form. This verification procedure, if true, makes the statement in (110) true. Therefore, we will call this type of verification procedure the ‘truth’ verification procedure. Notice, also that (111) contains a universal quantifier. Therefore, a universal quantifier is used in the verification procedure (*check all snails*), which distinguishes it from the next type of verification procedure we will present. Therefore, we will call this verification procedure the ‘universal’ verification procedure.

$$(112) \quad \forall x[\neg \text{snail}(x) \vee \text{antennae}(x)]$$

Notice that the statement in (112) is logically equivalent to the statement in (111). If true, the sentence in (110) will be true, and it contains a universal quantifier, leaving it identical in every respect we have covered. Therefore, we will continue referring to the form in (111).

There is more than one way to verify a sentence. A second way to verify the sentence in (110) is to look for information that would disconfirm the sentence, or show that the sentence is false. In order to do this, we can simply negate the statement in (111). If that statement is true, then we know that (110) is false. Let us go through the calculation of how we could arrive at this logical statement.

$$\neg \forall x[\text{snail}(x) \rightarrow \text{antennae}(x)] \tag{6.1}$$

¹Throughout this section, we will be talking about verification ‘procedures’. However, we will not be making claims as to the psychological reality of the steps we suggest. We will be presenting logical statements, and make claims about which are easier to verify. From this, we assume that the procedure used to carry out the harder logical form is ‘harder’ than the procedure used to carry out the easier logical form, without specifying the nature of these procedures.

$$\neg\forall x[\neg\text{snail}(x) \vee \text{antennae}(x)] \quad (6.2)$$

$$\exists x\neg[\neg\text{snail}(x) \vee \text{antennae}(x)] \quad (6.3)$$

$$\exists x[\text{snail}(x) \wedge \neg\text{antennae}(x)] \quad (6.4)$$

We begin with (111) as the statement the participant is supposed to verify. Equation (6.1) is (111) negated. We must rewrite the implication in terms of $\neg$ and $\vee$, as shown in Equation (6.2), and distribute this negation to get an existential, as in Equation (6.3). Next, we distribute the negation for the inner clause, as in Equation (6.4). This yields the statement that if verified will make (111) false. Paraphrased, the statement in Equation (6.4) states that *if you find a snail without antennae, then you can conclude that (110) is false.*² We will call this procedure (the paraphrase we have outlined) a ‘falsity’ verification procedure, because the criteria one is looking for is what determines that the utterance is false. Additionally, Equation (6.4) is different from (111) in that it contains an existential quantifier. Therefore, we will call the procedure associated with this equation an ‘existential’ verification procedure.

Given that every quantified sentence has both an existential and a universal verification procedure, it does not appear that adults implement them equally well. Wason (1965) showed, using Wason’s selection task, that adults have difficulty implementing the existential verification procedure. In this task, participants are shown a series of 4 cards, which have either letters or numbers written on each side of the card. Participants are told a sentence like (113), and their task is to determine which cards to flip over to verify if the sentence is true or false.

(113) If a card has a vowel on one side, then the other side has an even number

²This is equivalent to the modus tollens inference: given $p \rightarrow q$, it is logically entailed that $\neg q \rightarrow \neg p$.

The sentence in (113) can be restated as (114), and has two verification procedures, as exemplified by the statements in Equations (6.5) and (6.6).

(114) Every card with a vowel on one side has an even number on the other side

$$\forall x[\text{vowel}(x) \rightarrow \text{even}(x)] \quad (6.5)$$

$$\exists x[\text{vowel}(x) \wedge \neg \text{even}(x)] \quad (6.6)$$

Equation (6.5) is the statement that requires the verification procedure that is both ‘truth’ and ‘universal’. Equation (6.6) is the statement that requires the verification procedure that is both ‘falsity’ and ‘existential’. If it were the case that all of the cards had vowels displayed, then the verification procedure would be trivially easy: that is to flip over each card that has a vowel on it to check if it had an even number on the other side, as in Equation (6.5). However, in the task, some cards have letters showing, and some cards have numbers showing, making it impossible to check only the cards with vowels (because one does not know which cards have vowels), as shown in Figure (6.1).



Figure 6.1: Wason’s Selection Task

In this situation, one must resort to the second verification procedure: that is to check the cards with vowels, and check cards that do not have an even number, that is the A and the 7. This is the verification procedure required by the statement in Equation (6.6), which is a falsity procedure. That is, if an adult finds an odd number that has a vowel on the other side, then statement (114) would be false.

Adults, however, frequently fail in this sort of task, turning over vowel cards and cards with an even number 4% of the time (Wason and Johnson-Laird, 1972). Wason and Johnson-Laird (1972) found that adults arrived at the correct answer only 4% of the time, but Evans (1982) reported variance on participants arriving at the correct answer 6-33% of the time.

From this experiment, it appears that when the first verification procedure is unavailable, adults struggle with the second verification procedure, leading us to conclude that there is something difficult about this second verification procedure.

To this point, there are two confounded factors that may each account for the difficulty associated with implementing the second verification procedure. One, it may be that adults do not like to look for disconfirming evidence (do not use falsification procedures). Two, it may be the case that existential verification procedures are more difficult than universal verification procedures. The current task only shows us that adults have difficulty with a verification procedure that is both ‘falsity’ and ‘existential’. We will now turn to scope ambiguity, where we argue that both of these notions are necessary for capturing the relevant facts.

6.2 Extended to Scope Ambiguities

Let us turn to an investigation of how these verification procedures apply to scope ambiguities. The experiments we have shown suggest that surface scope interpretations gain an advantage over inverse scope interpretations, and we have shown that this advantage cannot be attributed to the parser. We are investigating verification procedures as a source of this advantage. But, why would the verification of the inverse scope interpretation be harder than for the surface scope interpretation? We claim

that verifying inverse scope interpretations is identical to the verification procedure that adults find difficult in Wason’s selection task.

As we have shown, a universally quantified sentence is logically identical to an entailment. Therefore, we will make the assumption here that their verification procedures are the same. Let us first look at the surface scope interpretation of a string like (115).

(115) Every snail doesn’t have antennae

The surface scope interpretation, *every snail is such that it is missing antennae*, can be written as in Equation (6.7).

$$\forall x[\text{snail}(x) \rightarrow \neg \text{antennae}(x)] \quad (6.7)$$

The truth verification procedure dictates that one *checks each snail in the relevant universe and makes sure it is antennae-less*. This is also a universal verification procedure. From this perspective, verifying surface scope interpretations is no different from verifying simple *every* statements. For every interpretation, there are two verification procedures. The ‘falsity’ verification procedure for the surface scope interpretation is $\exists x[\text{snail}(x) \wedge \text{antennae}(x)]$, just as in the simple *every* statements, and it is an existential. Because (6.7) is simply (111) with the *antennae* predicate replaced with $\neg \text{antennae}$, we predict the truth (and universal) verification procedure to be easy to verify and the falsity verification procedure (and existential) to be difficult to verify.

Let us turn to the inverse scope interpretation of (115), which is *it is not the case that every snail has antennae*. The logical description of this interpretation is shown in Equation (6.8).

$$\neg\forall x[\text{snail}(x) \rightarrow \text{antennae}(x)] \quad (6.8)$$

$$\exists x[\text{snail}(x) \wedge \neg\text{antennae}(x)] \quad (6.9)$$

$$\neg\neg\forall x[\text{snail}(x) \rightarrow \text{antennae}(x)] \quad (6.10)$$

Again, we have two possible verification procedures, which we will outline here. The truth verification procedure is represented by the statement in Equation (6.9). The falsity verification procedure is represented by the statement in (6.10). Notice here, that the truth verification procedure corresponds to the existential verification procedure, and the falsity verification procedure is the universal verification procedure.

Recall, from Wason's selection task, we found that adults failed to execute the 'falsity' 'existential' verification procedure, but we did not know which of these features to attribute the difficulty to. If it is the case that the existential verification procedure is the only aspect that is difficult, then we expect adults to easily implement the universal verification procedure. If it is the case that falsity verification procedure is the only thing that is difficult, then we would expect adults to easily execute the truth procedure here. If it is the case that both factors play a role, then we may expect some degree of inability to use either verification procedure. Let us turn to our results from the IVT, which bear on this issue.

We will first determine whether adults use the universal verification procedure for inverse scope interpretations. Paraphrased, this verification procedure is an instruction to *check every snail, and see whether it has antennae*. If this is true, the statement (109) is false. It does not appear that adults, when accessing the inverse scope interpretation, use this verification procedure. That is, if they did, we would

have seen adults, in the IVT, persisting to the last cup, but giving a T/F response that corresponded to the inverse scope interpretation, which did not occur. Therefore, we can conclude that adults do not seem to carry out ‘falsity’ verification procedures.

Let us see how adults carry out the existential verification procedure. If it is the case that the existential contributes no difficulty, this would mean adults would look for an antennae-less thing that is a snail with no additional difficulty. In the IVT, we find that adults are capable of performing this procedure, as they obtain inverse scope in the IVT about 30% of the time. However, it does not appear to be the case that this interpretation is as equally accessible as the universal in the surface scope case.

We propose that the best explanation for the IVT facts is that both existential and falsity verification procedures come with some degree of difficulty. These are confounded in the Wason’s selection task and surface scope interpretations, but with inverse scope interpretations, we are able to view this relative difficulty. It seems clear that falsity procedures are rarely pursued, regardless of interpretation. It also seems clear that truth procedures aren’t created equally. We find a correlation between the interpretation adults prefer in the IVT and the universal verification procedure. From this correlation, we claim that the inverse scope interpretation is dispreferred because it utilizes the existential verification procedure, which requires the distribution of sentential negation. This is dispreferred with respect to the universal verification procedure.

Let us put forth a beginning hypothesis. The grammar, when it encounters a scopally ambiguous string, makes available both interpretations. The language processing system then attempts to verify both possible interpretations. In a sense, it is a race to see which can be verified first. When a scopally ambiguous string is heard,

there are two interpretations and two truth verification procedures. Because the surface scope interpretation is easier to verify, the surface scope interpretation gains an advantage, and because it is the first that can be verified, it is the first interpretation selected. Therefore, the advantage for surface scope interpretations falls out of a competition fact about their verification procedures. Of course, the advantage for the surface scope verification procedure is simply one of many post-parsing factors that contribute to final interpretation. Because the inverse scope interpretation is readily utilized in conversation, we must acknowledge that other factors work to outweigh this particular factor.

6.3 The Existential Verification Procedure

To this point, we have posited that adults do not carry out falsity procedures, and that existential verification procedures are disadvantaged with respect to universal verification procedures. Here, we review more evidence that the existential verification procedure is conducted by adults, but that it is only dispreferred relative to a quicker verification procedure. Recall that in the IVT, adults performed quite well with sentences like (116).

(116) Not every snail has antennae

The reason these sentences were informative is because their logical form is the exact same as the inverse scope interpretation's logical form. Because their logical forms are exactly the same, so are their verification procedures. The logical form is shown in Equation (6.11), with the truth verification statement (Equation (6.12)) and falsity verification statement (Equation (6.13)).

$$\neg\forall x[\text{snail}(x) \rightarrow \text{antennae}(x)] \quad (6.11)$$

$$\exists x[\text{snail}(x) \wedge \neg \text{antennae}(x)] \quad (6.12)$$

$$\neg \neg \forall x[\text{snail}(x) \rightarrow \text{antennae}(x)] \quad (6.13)$$

Here, too, we know that adults, when presented with (116) do not use the falsity verification procedure, as they are able to stop on the first cup.

Adults performed well on this task. That is, they stopped on the first cup, and gave an appropriate truth condition, suggesting they were using the verification procedure represented in Equation (6.12). This suggests that the existential is possible to utilize, and the difficulty with the existential only arises when there is a quicker verification procedure tied to an alternate interpretation. For the string in (116), there is no alternative interpretation.

Therefore, we conclude that existentials are harder than universals, when they are competing as alternative interpretations of a single sentence. This captures the observation that inverse scope interpretations are harder to verify than surface scope interpretations, but that the existential verification procedure is utilized when there are no competing interpretations.

6.4 Reflecting on our Previous Findings

We have claimed that the advantage gained by surface scope interpretations is due to a faster verification procedure than is available for inverse scope interpretations. We will now turn to the ramifications of this claim to the findings we have presented in this dissertation. We show that this claim even naturally captures some findings that, until now, remained mysterious.

6.4.1 S-structure vs. LF Ambiguity

We have suggested that surface scope interpretations gain an advantage over inverse scope interpretations due to the ease of their verification procedures. Therefore, when no other factors are present to influence interpretation, surface scope interpretations will be selected.

At the conclusion of chapter 2, we concluded that because adults could not access both interpretations of scopally ambiguous strings, that LF ambiguity was processed serially, at least before the situational disambiguation was reached. Our hypothesis, that verification can lend an advantage to surface scope interpretations, allows for the early disambiguation we observed to take place. Additionally, we observed this seriality was not present in other sorts of ambiguity, such as PP attachment ambiguity. The reason that PP attachment ambiguity maintained its parallel appearance in the IVT is because the two interpretations utilize the same type of verification procedure. Therefore, in that domain, neither obtains an advantage, and there is no early disambiguation.

6.4.2 On Previous Results

Anderson (2003) showed that for strings like (117) and (118), adults adhere to surface scope interpretations in online processing.

(117) A climber scaled every cliff

(118) Every climber scaled a cliff

The logical form for the surface scope interpretation of (117) (and the inverse scope interpretation for (118) is shown in (119). The logical form for the inverse scope interpretation of (117) (and the surface scope interpretation for (118) is shown in (120).

(119) $\exists x[\text{climber}(x) \wedge \forall y[\text{cliff}(y) \rightarrow \text{scaled}(x, y)]]$

(120) $\forall x[\text{cliff}(x) \rightarrow \exists y[\text{climber}(y) \rightarrow \text{scaled}(y, x)]]$

For sentences like (118), our account correctly predicts that surface scope interpretations would be easier to verify than inverse scope interpretations, yielding the ease of processing for surface scope interpretations. This is because the surface scope interpretation of (118), shown in (120), utilizes a universal verification procedure, while the inverse scope interpretation, shown in (119), requires use of the existential verification procedure.

However, because the logical forms for the two interpretations of (117) and (118) are identical, we should predict the opposite for the processing of (117). That is, the the surface scope interpretation of (117), shown in (119), utilizes an existential verification procedure, while the inverse scope interpretation, shown in (120), requires use of a universal verification procedure. This predicts that in this case the inverse scope interpretation should be easier to verify, a prediction that is not borne out by Anderson's processing results. Therefore, we must explain why this sentence behaves differently than predicted. We desire to keep our account consistent across all types of semantic ambiguity, so the existence of an exception would be quite undesirable.

We suggest that strings like (117) exhibit a surface scope preference for reasons independent of processing. (117) contains a subject indefinite, which in an experimental context is likely to be interpreted as specific, that is *there exists a climber*. Once it is the case that the subject indefinite is treated specifically, then it will be interpreted as a topic, and take wide scope over the object. This account predicts that sentences like (117) will almost always be interpreted with surface scope interpretations, not just in online processing. The first experiment from Anderson's work bears out this prediction. Anderson gives participants sentences like (117), and asks them to select

whether there exists a single climber or multiple climbers. In this case, participants select a *single climber* 81% of the time. This demonstrates an extremely strong bias for surface scope interpretations, independent of online processing.

We suggest that Anderson’s results can be accounted for by a combination of our account, and treating specific subject indefinites as taking wide scope. This allows us to extend our account to more types of scope ambiguity, suggesting that we are looking at a unary phenomenon for semantic ambiguity resolution.

6.4.3 IVT: Number

Recall, the results from IVT: number did not fit in seamlessly with IVT: universal quantifier. We investigated sentences like (121).

(121) Jon didn’t find two hearts

There were two findings. First, the condition (which interpretation could be verified first) had no influence on interpretation. These results fit with the results from those of other IVT experiments in that it appears that adults select an interpretation before encountering disambiguating evidence. This seems to be a general property of the ambiguity resolution process for scopally ambiguous strings. Second, we found that adults adhered to the surface scope interpretation 43.75% of the time (averaged across conditions). This finding appears to contrast with the findings from other IVT experiments, in that there is no strong preference for surface scope interpretations. Now that we have outlined in more detail the factors underlying the selection of an interpretation, let us look at these results. In this section, we suggest that there does not exist a verification procedure that is easier than the procedure for the other interpretation, yielding no overall preference for interpretation.

(121) has two interpretations: the surface scope interpretation, *it is not the case there exist two that were found*, shown in Equation (6.14) and the inverse scope interpretation, *two are such that there are not found*, shown in Equation (6.15).

$$\neg \exists x, \exists y (x \neq y) [\text{heart}(x) \wedge \text{found}(x)] \wedge [\text{heart}(y) \wedge \text{found}(y)] \quad (6.14)$$

$$\exists x, \exists y (x \neq y) [\text{heart}(x) \wedge \neg \text{found}(x)] \wedge [\text{heart}(y) \wedge \neg \text{found}(y)] \quad (6.15)$$

Because we have determined that adults utilize only truth verification procedures (and not falsity verification procedures), let us write out the truth verification procedure for each. First, the truth verification procedure for the surface scope interpretation is presented in Equation (6.16).

$$\forall x, \forall y (x = y) \vee [\neg \text{heart}(x) \vee \neg \text{found}(x)] \vee [\neg \text{heart}(y) \vee \neg \text{found}(y)] \quad (6.16)$$

Paraphrased, the verification procedure in Equation (6.16) says to *look at all items, and there will not be two that are hearts that have the property of being found*.

The verification procedure for the inverse scope interpretation is presented in Equation (6.17).

$$\exists x, \exists y (x \neq y) [\text{heart}(x) \wedge \neg \text{found}(x)] \wedge [\text{heart}(y) \wedge \neg \text{found}(y)] \quad (6.17)$$

Paraphrased, this verification procedure says to *look for two items that are hearts that were not found*.

In the last section, we concluded that existential verification procedures are harder to utilize than universal verification procedures. We concluded this on the basis of the

observation that inverse scope interpretations were harder to access than surface scope interpretations for sentences containing the subject universal quantifier and negation. If this conclusion is accurate, then we would predict that in the case of negation and a numerically quantified object, the surface scope interpretation would be easier to verify, and thus, we should observe mostly surface scope interpretations in the IVT. This prediction is not borne out, as we observed 43% surface scope interpretations, suggesting that our previous analysis requires adjustment.

The surface scope interpretation of the strings containing negation and a numerically quantified object does have one property in common with the inverse scope interpretation of sentences containing negation and a universally quantified subject. Both interpretations require the distribution of sentential negation to compute the truth verification procedure. Here, we posit that the distribution of sentential negation makes a verification procedure more difficult. Let us see how this revision bears on our previous results.

Concerning strings containing a universally quantified subject and negation, the truth verification procedure for the inverse scope interpretation requires the distribution of sentential negation, and utilizes an existential verification procedure. These two factors, together, make the inverse scope interpretation more difficult to verify than the surface scope interpretation, which utilizes a universal verification procedure and does not require the distribution of sentential negation. Therefore, even though the parser makes both interpretations available, the ease of the verification procedure associated with surface scope interpretations makes adults select surface scope interpretations more easily than inverse scope interpretations.

Now, let us turn to strings containing negation and a numerically quantified object. For these sentences, the surface scope interpretation requires the distribution of

negation, and the inverse scope interpretation utilizes an existential verification procedure. Therefore, it does not appear that either interpretation gains an advantage due to ease of verification procedure. When the parser makes both interpretations equally available, there is no effect of verification procedures, and adults select the surface scope interpretation at about chance.

An interesting question arises when we look at the responses for IVT: Number more closely. 9/20 adults obtained the surface scope interpretation on all (or all but one) trials, and 5/20 adults obtained the inverse scope interpretation on all (or all but one) trials, showing a bimodal distribution. This bimodal distribution may represent two distinct phenomena. One, it may be the case that we are observing the effects of priming. That is, an adult may obtain either interpretation on the first trial, and the majority of participants continue to adhere to the interpretation they obtained on the first trial. However, it may also be the case that adults attribute different amounts of difficulty to different verification procedures. The bimodal distribution we observe may be accounted for by some adults preferring to carry out existential verification procedures, and others preferring to distribute negation, but that these factors were confounded in the case of sentences containing the universal quantifier. Further research is needed to differentiate between these two accounts.

A second question arises concerning this experiment. In the case of IVT: universal quantifier, it was clear that because an easy verification procedure forced adults to select a surface scope interpretation, adults ignored information in the situation and always pursued that interpretation. However, in the case of IVT: number, adults seem, in principle, capable of accessing either interpretation, but they cannot alter this interpretation as a function of information in the situation. That is, the results from this experiment did not differ as a function of which interpretation could be verified

first. Therefore, it seems that for scope ambiguity, adults select an interpretation, although in this case, it is not clear why. These findings differ from those of IVT: syntactic ambiguity, where adults appear able to maintain multiple interpretations, and can use information in the scene to guide their choice of interpretation. These findings raise the possibility that the difference between s-structure and LF ambiguity lies in the ease of revision available for the two types of constructions. For example, these findings would be explained if, for some reason, adults can revise in the case of s-structure ambiguity based on information in the scene, but not for LF ambiguity. If this is the case, we can ask if it is the case that the cues for revision differ in the two cases. Although this is an interesting area for future research, our tentative conclusion is that it seems clear that there exists some difference in the processing of s-structure and LF ambiguity, and that the choices made in the case of LF ambiguity can be described by appealing to the differing ease of the available verification procedures.

In summary, the IVT: number results forced us to modify our understanding of the relative difficulty of verification procedures, but supported the claim that the parser, in cases of scope ambiguity, outputs two interpretations without assigning preference to either interpretation.

6.4.4 Priming

We have suggested that surface scope interpretations gain an advantage via verification procedures, and that this advantage surfaces as a preference for surface scope interpretations when no other interpretive factors intervene. Priming is a factor that can intervene. Here, we will review a few priming results that remained mysterious until this point, and suggest that some of the priming effects we have observed are actually procedures that alleviate the difficulty associated with the inverse scope verification procedure.

The first observation about priming (in children and with the IVT) was that a preceding sentence made the inverse scope interpretation easier to access. Musolino and Lidz (2006) show that children obtain higher rates of inverse scope interpretations in (123) than in (122).

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

(123) Every horse jumped over the rock, but every horse didn't jump over the fence

This increase in access to inverse scope interpretations is explained by the fact that the positive lead-in sentence makes negation more felicitous.

Negative statements are easier to process when certain felicity conditions are met. Psychologically, negation seems to require an affirmative sentence against which to operate (Wason, 1965)

That is, the positive lead-in sentence sets up an expectation, which the negation can then negate. This appears to be a general property about negation, as Wason (1965) showed that adults could more quickly judge the truth of (125) over (124).

(124) 5 is not an even number

(125) 4 is an even number, but 5 is not an even number

Again, the negative statement is more felicitously processed after the presence of an affirmative statement which it can negate.

Although positive statements aid the processing of negation, it is also the case that the benefit attributed to (123) over (122) can also be described by improving the richness of the discourse. It has been shown that adult's ability to calculate the existential verification procedure can be assisted by a discourse environment. Performance in Wason's selection task, which requires the usage of the existential

verification procedure, is greatly improved when the items involved are not numbers, but are real-life situations. For example, if participants are told to verify the statement in (126), correct performance reaches 70% (Griggs and Cox, 1982).

(126) If a person has a beer, they must be 21 years old

That is, participants correctly observe that they should check all persons with a beer, and all persons under the age of 21. Participants no longer say that they should check all of those over the age of 21, which is the error that was made when the participants were simply presented with the cards with vowels and numbers. Therefore, we can conclude that an appropriate context allows adults to better access their existential verification procedure. For scope ambiguity, it seems that a richer discourse (a positive lead-in sentence) nullifies the advantage gained by the surface scope verification procedure, thereby increasing rates of obtaining the inverse scope interpretation.

A second mystery in the domain of priming is why a *not every* sentence would prime the inverse scope interpretation of a *every x didn't* sentence, given that these two sentences share nothing in terms of syntactic form. That is, we seem forced into concluding that priming is unrelated to a relation between PFs and LFs, but is an activated logical form representation. However, we could also view priming as another way to alleviate the difficulty associated with the inverse scope verification procedure. Let us formulate 'verification procedure priming' as a pressure to use a verification procedure that has been recently used. When a *not every* sentence is verified, the existential verification procedure is necessarily used. Next, when an *every x didn't* sentence is encountered, there exists a bias to use the same verification procedure that has recently been used, which leads to the inverse scope interpretation. That is, executing the existential verification procedure (which is easy in a *not every* sentence because there is no competing procedure), temporarily makes this procedure easier to

implement even when there is competition involved. We are not attempting to argue against logical form priming, but present the ‘priming’ of verification procedures as an alternative explanation. More research on semantic priming most likely will reveal which sort of priming is most natural.

We have not provided a complete account of LF priming, and there is much research left to be conducted. However, we have provided an initial account of the effects observed in this dissertation (and previous research) regarding the priming of interpretations of scopally ambiguous strings. We have suggested that many of these observations can be accounted for by appealing to the priming of verification procedures.

6.4.5 On Children

We have suggested that the inverse scope interpretation is more difficult to verify than the surface scope interpretation, and this is the source of the advantage for surface scope interpretations in adults. On the surface, this appears to be a poor account for the child facts. That is, if inverse scope is so difficult to verify, how is it the case that four year olds can do it (and even prefer to do so)? And how is the difficulty of verification related to the integration of discourse information?

Recall that we have claimed that surface scope interpretations gain an advantage only in a race model. That is, when the parser makes both interpretations equally available, and there are no additional extra-linguistic factors, then surface scope interpretations gain an advantage. However, recall that our model suggested that for four year olds, it is not the case that both interpretations are made equally available by the parser. If it is the case that four year olds encode a bias for inverse scope as a parsing preference, then there will be no ‘race’ of verification procedures.

Surface scope interpretation-children, who happen to be five years old, are adult-like in that they have parsers that make both interpretations equally available. Here, then, we see the advantage for surface scope interpretations reappear. Five year olds are different from adults in that, in a TVJT, they do not seem to make use of other extra-linguistic factors that, for adults, override this advantage for surface scope interpretations. Therefore, the verification advantage for surface scope interpretations fits naturally with the child data we have reviewed.

6.5 Conclusions on Verification

In this section, we outlined a hypothesis that verification procedures are responsible for the advantage for surface scope interpretations. We suggest that when the parser makes two interpretations equally available, the interpretation that can be verified the quickest will gain an advantage. Therefore in tasks in which speed is required (speeded forced choice), or when there are no other extra-linguistic forces (IVT), surface scope interpretations are selected. This advantage can be overridden by other discourse factors (in the TVJT) or by a parser bias (as in the case of four year olds).

The idea that one verification procedure is preferred over another was supported by findings from the psychology of reasoning, which suggests that not all verifications strategies are created equal (Wason, 1965). Additionally, we showed that the same manipulations that assist adults with solving Wason's selection task also help in accessing inverse scope interpretations, suggesting a similar mechanism underlies both.

The claim that different verification procedures decide between two equally available interpretations fits nicely with both our s-structure ambiguity findings and our

findings with children. We now know that syntactic ambiguity is different from scope ambiguity because the two interpretations of a s-structurally ambiguous sentence do not appeal to different types of verification procedures, which in turn means that they remain equally available. Additionally, our model suggested that four year olds were able to access inverse scope interpretations (without access to discourse information) via a parsing bias, which then obviates the role of verification procedures in deciding interpretation. When five year olds eliminate this parsing bias, we are then able to account for the strong advantage of surface scope interpretations. Under this view, the IVT in adults is strikingly similar to 5 year old children because they are both cases where the advantage for surface scope takes hold, without interference from later discourse processes.

To conclude, we have a prediction concerning verification and working memory that we must leave for future research. We found that interpretation was not affected by manipulations that taxed working memory. This is because we used a task which did not require verification. Therefore, the manipulation did little to impact the parser. However, we predict that these manipulations would affect interpretation in a task in which verification is required. More generally, we have presented a novel account of the advantage for surface scope interpretations.

Chapter 7

Conclusions

The main focus of this dissertation was to investigate processes involved in LF parsing. To accomplish this investigation, we looked at one construction in detail over various developmental timepoints and in different experimental situations to identify the factors that contribute to final interpretation. In doing this, we were able to identify the contribution of parsing biases and extra-linguistic information in the processing of scopally ambiguous strings. This investigation outlines a framework for the investigation of other types of ambiguity, and particularly, focuses on the necessity of considering developmental and task factors in identifying the processes involved in adult sentence processing.

In this dissertation, we investigated scopally ambiguous strings, particularly those with a universally quantified subject and negation, as in (127).

(127) Everyone doesn't like eating in large groups of people

Strings that permit multiple interpretations of one s-structure representation allow us to identify how the parser selects from multiple available analyses. With respect to this construction in particular, it had previously been observed (from both the child

language and adult processing literature) that surface scope interpretations gained an advantage over inverse scope interpretations. By testing this advantage under different experimental situations, and over different developmental timepoints, we gained insight into how the sentence processing mechanism operates, particularly with respect to the understudied process of LF parsing.

This investigation served to identify the nature of LF parsing, and focused on two main questions concerning the resolution of scopal ambiguity. First, what accounts for the advantage for surface scope interpretations in adults? Previous research and our experiments from section 2 showed that in certain situations, adults obtain largely surface scope interpretations. Second, what describes the advantage for surface scope interpretations in children? Previous research showed that children frequently adhere to surface scope interpretations, even when adults accessed inverse scope interpretations (Musolino et al., 2000; Musolino and Lidz, 2006). In both the children and adults, the advantage for surface scope interpretations could be described by either a parsing default (Anderson, 2003; Musolino and Lidz, 2005), or by extra-linguistic factors. We pursued the view that these two findings were related, using information from development to identify the contribution of the parser to interpretation.

Concerning the source of the surface scope advantage in adults, in section 2, we showed that a preference for surface scope interpretations holds in the IVT and also in the speeded forced choice task, confirming the finding that has been reported with other types of scope ambiguity (Tunstall, 1998; Anderson, 2003). In section 5.1, we suggested that the advantage for surface scope interpretations cannot arise from a parsing bias in the adult system. In section 6, we suggest that surface scope interpretations are easier to verify, giving surface scope interpretations an advantage. Therefore, we conclude that there is no linguistic preference given to surface scope

interpretations. The parser makes available two interpretations for one PF, and in cases where no additional information is utilized, the easiest interpretation to verify will be pursued.

Concerning the surface scope advantage in children, in section 4.1, we revisited the observation of isomorphism. We found that, although the observation of isomorphism holds, it only represents a limited window of development. We found that 4 year olds readily accessed inverse scope interpretations, at adult-like levels. This shows that it cannot straight-forwardly be the case that children lack the resources to access inverse scope interpretations. We claim that young children are able to obtain inverse scope interpretations via a parsing bias, and when the older children acquire adult-like parsers, this results in an increase in surface structure interpretations. This model of U-shaped development was supported by the finding that children do not utilize discourse information to guide their selection of an interpretation, as shown in section 4.2.3. In section 5.1, we showed that our U-shaped model could be analyzed by the same model used to account for the adult results. This investigation into the role of development, parsing biases and discourse biases under a unified framework allowed us to identify the exact role of the parser in the resolution of LF ambiguity, a process which has not been studied in detail.

The main goal of psycholinguistic research is to identify how the human sentence processing mechanism operates, that is, how a speaker gets from a string to an interpretation. Although much research has been conducted on how a speaker obtains an s-structure, very little work has investigated how a speaker accesses a logical form. However, it is crucial to understand how one accesses a logical form, if one desires a complete description of the human sentence processing mechanism. In this dissertation, we contribute to the significant amount of research in sentence processing

(which has primarily focused on s-structure parsing) by presenting a detailed study of the role of the parser in how speakers obtain a logical form. We claim that investigating data from children and adults in tandem is crucial for investigating this process. We show that data from children's understanding of ambiguous string illuminates our understanding of how adults process these sentences, and furthermore, that data from adults is necessary for understanding how it is that children's parsers operate. Therefore, in studying children and adults together, we are able to provide a complete picture of the procedures that underlie LF parsing.

Bibliography

- Anderson, C. (2003). The Structure and Real-Time Comprehension of Quantifier Scope Ambiguity. PhD thesis, Northwestern University.
- Baddeley, A. (1986). Working Memory. Oxford University Press, Oxford.
- Baddeley, A. and Hitch, G. (1974). Working memory. In Bower, G., editor, The Psychology of Learning and Motivation, pages 47–90. Academic Press, New York.
- Baddeley, A., Lewis, V., Eldridge, M., and Thomson, N. (1984). Attention and retrieval from long term memory. Journal of Experimental Psychology: General, 113:518–530.
- Baddeley, A., Vallar, G., and Wilson, B. (1987). Sentence comprehension and phonological memory: some neuropsychological evidence. In Coltheart, M., editor, Attention and performance XII, pages 509–529. Laurence Erlbaum Associates, Hillsdale, NJ.
- Beghelli, F. (1995). The phrase structure of quantifier scope. PhD thesis, University of Massachusetts.
- Bever, T. (1970). The cognitive basis for linguistic structures. In J.R.Hayes, editor, Cognition and the Development of Language, pages 227–352. Wiley, New York, NY.

- Boland, J. and Blodgett, A. (2001). Understanding the constraints on syntactic generation: Lexical bias and discourse congruency effects on eye movements. Journal of Memory and Language, 45:391–411.
- Braine, M. and O'Brien, D. (1998). Mental Logic. Lawrence Erlbaum Associates, Mahwah, New Jersey.
- Bransford, J., Barclay, J., and Franks, J. (1972). Sentence memory: A constructive versus interpretive approach. Cognitive Psychology, 3:193–209.
- Brown, R. (1972). A First Language: the early stages. Harvard University Press, Cambridge, MA.
- Carlson, K. (2002). Parallelism and Prosody in the Processing of Ellipsis. Routledge, New York and London.
- Carlson, K., Dickey, M., and Kennedy, C. (2005). Structural economy in the processing and representation of gapping sentences. Syntax, 8:208–228.
- Carpenter, P., Miyake, A., and Just, M. (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia and aging. In Gernsbacher, M., editor, Handbook of Psycholinguistics, pages 1075–1122. Academic Press, San Diego, CA.
- Clifton, C. and Ferreira, F. (1989). Ambiguity in context. Language and Cognitive Processes, 4:77–103.
- Colle, H. and Welsh, A. (1976). Acoustic masking in primary memory. Journal of Verbal Learning and Verbal Behavior, 15:17–32.
- Cowan, N. (1999). An embedded-process model of working memory. In Models of Working Memory, pages 62–101. Cambridge University Press, New York, NY.

- Crain, S. and Steedman, M. (1985). On not being led up the garden path: the use of context by the psychological syntax processor. In Dowty, D., Karttunen, L., and zwicky, A., editors, Natural Language Parsing, page 320. Cambridge University Press, Cambridge, MA.
- Crain, S. and Thornton, R. (1998). Investigations in Universal Grammar: A guide to Research on the Acquisition of Syntax and Semantics. MIT Press, Cambridge, MA.
- Diesing, M. (1992). Indefinites. MIT Press, Cambridge, MA.
- Estes, W. (1973). Phonemic coding and rehearsal in short-term memory. Journal of Verbal Learning and Verbal Behavior, 12:360–372.
- Evans, J. S. (1982). The Psychology of Deductive Reasoning. Routledge and Kegan Paul.
- Ferreira, V. (2003). The processing basis of syntactic persistence: we repeat what we learn. Presented at XIII Conference of the European Society of Cognitive Psychology, Granada.
- Fodor, J. (2002). Psycholinguistics cannot escape prosody.
- Fox, D. (2000). Economy and Semantic Derivation. MIT Press.
- Frazier, L. and Fodor, J. (1978). The sausage machine: a new two-stage parsing model. Cognition, 6:291–325.
- Frazier, L. and Rayner, K. (1982). Making and correcting errors during sentence comprehension: eye movements in the analysis of structurally ambiguous sentences. Cognitive Psychology, 14:178–210.
- Gil, D. (1982). Quantifier scope and linguistic variation. Linguistics and Philosophy, 5:421–472.

- Griggs, R. and Cox, J. (1982). The elusive thematic-materials effect in wason's selection task. British Journal of Psychology, 73:407-420.
- Grodner, D., Gibson, E., and Watson, D. (2005). The influence of contextual contrast on syntactic processing: Evidence for strong-interaction in sentence comprehension. Cognition, 95:275-296.
- Grodzinsky, Y. and Reinhart, T. (1993). The innateness of binding and coreference. Linguistic Inquiry, 29:69-101.
- Gualmini, A. (2004). Some knowledge children don't lack. Linguistics, 42:957-982.
- Gualmini, A., Hacquard, V., Husley, S., and Fox, D. (2005). Beyond the surface of child language. Presented at LSA Annual Meeting.
- Heim, I. and Kratzer, A. (1998). Semantics in Generative Grammar. Blackwell.
- Horn, L. (1989). A Natural History of Negation. University of Chicago Press, Chicago.
- Hornstein, N. (1995). Logical Form: From GB to Minimalism. Blackwell, Oxford.
- Hsu, C.-C. (2006). Issues in Head-Final Relative Clauses in Chinese: Derivation, Processing and Acquisition. PhD thesis, University of Delaware.
- Hulsey, S., Hacquard, V., Fox, D., and Gualmini, A. (2004). The question answer requirement and scope assignment. In Plato's Problems: Papers on Language Acquisition. MITWPL 48. MITWPL, Cambridge, MA.
- Jackendoff, R. (1972). Semantic Interpretation in Generative Grammar. MIT Press, Cambridge, MA.
- Johnson-Laird, P. (1975). Models of deduction. In Falmange, R., editor, Reasoning: Representation and Process in Children and Adults. Lawrence Erlbaum Associates, Hillsdale, NJ.

- Kimball, J. (1973). Seven principles of surface structure parsing in natural language. Cognition, 2:15–47.
- King, J. and Just, M. (1991). Individual differences in syntactic processing: the role of working memory. Journal of Memory and Language, 30:580–602.
- Kurtzman, H. and MacDonald, M. (1993). Resolution of quantifier scope ambiguities. Cognition, 48:243–279.
- Leddon, E., Lidz, J., and Pierrehumbert, J. (2003). Suprasegmental cues to meaning in child-directed speech. manuscript.
- Lidz, J. and Conroy, A. (2007). Mechanisms of If priming: Data from kannada and english. Presented at the Boston University Conference on Language Development.
- Lidz, J., McMahon, E., Syrett, K., Viau, J., Anggoro, F., Peterson-Hicks, J., Sneed, E., Bunger, A., Flevaris, T., Graham, A., Grohne, K., Lee, Y., and Strid, J. (2004). Quantifier raising in 4-year-olds. Proceedings of 28th Boston University Conference on Language Development, pages 340–349.
- Lidz, J. and Musolino, J. (2002). Children’s command of quantification. Cognition, 84:113–154.
- Macdonald, M., Just, M., and Carpenter, P. (1992). Working memory constraints on the processing of syntactic ambiguity. Cognitive Psychology, 24:56–98.
- MacDonald, M., Pearlmutter, N., and Seidenberg, M. (1994). The lexical nature of syntactic ambiguity. Cognitive Psychology, 101:676–703.
- Martin, R., Wogalter, M., and Forlano, J. (1988). Reading comprehension in the presence of unattended speech and music. Journal of Memory and Language, 27:382–398.

- McRae, K., Spivey-knowlton, M., and Tanenhaus, M. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. Journal of Memory and Language, 38:283–312.
- Musolino, J., Crain, S., and Thornton, R. (2000). Navigating negative quantificational space. Linguistics, 38:1–32.
- Musolino, J. and Lidz, J. (2003). Turning adults into kids. Language Acquisition, 11:277–291.
- Musolino, J. and Lidz, J. (2005). On the quantificational status of indefinites: The view from child language. Language Acquisition, 13:73–102.
- Musolino, J. and Lidz, J. (2006). Why children aren't universally successful with quantification. Linguistics, 44:817–852.
- Noveck, I. (2001). When children are more logical than adults: experimental investigations of scalar implicature. Cognition, 78:165–188.
- Novick, J., Trueswell, J., and Thompson-Schill, S. (2005). Cognitive control and parsing: Re-examining the role of broca's area in sentence comprehension. Journal of Cognitive, Affective, and Behavioral Neuroscience, 5:263–281.
- Pickering, M. and vanGompel, R. (2006). Syntactic parsing. In Traxler, M. and Gernbacher, M., editors, Handbook of Psycholinguistics, pages 455–503. Elsevier, Amsterdam.
- Saffran, E. and Martin, O. (1975). Immediate memory for word lists and sentences in a patient with deficient auditory short-term memory. Brain and Language, 2:420–433.
- Scarborough, D. (1977). Frequency and repetition effects in lexical memory. Journal of experimental psychology: human perception and performance, 3:1–17.

- Steedman, M. (2001). The Syntactic Process. MIT Press, Cambridge, MA.
- Sturt, P., Pickering, M., and Crocker, M. (2000). Search strategies in syntactic re-analysis. Journal of Psycholinguistic Research, 29:183–194.
- Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., and Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. Science, 268:1632–1624.
- Thornton, R. and Wexler, K. (1999). Principle B, VP ellipsis and Interpretation in child grammar. MIT Press, Cambridge, MA.
- Trueswell, J. (1996). The role of lexical frequency in syntactic ambiguity resolution. Journal of Memory and Language, 35:566–585.
- Trueswell, J., Sekerina, I., Hilland, N., and Logrip, M. (1999). The kindergarden-path effect: Studying online processing in young children. Cognition, 73:89–134.
- Tunstall, S. (1998). The interpretation of quantifiers: Semantics and Processing. PhD thesis, University of Massachusetts.
- Viau, J., Musolino, J., and Lidz, J. (2006). Priming of abstract logical representations in 4-year olds. presented at the 30th Boston University Conference on Language Development.
- Wason, P. (1965). The context of plausible denial. Journal of Verbal Learning and Behavior, 4:7–11.
- Wason, P. and Johnson-Laird, P. (1972). Psychology of Reasoning: Structure and Content. Harvard University Press.
- Waters, G., Caplan, D., and Hildebrandt, N. (1987). Working memory and written sentence comprehension. In Coltheart, M., editor, Attention and performance XII, pages 531–555. Laurence Erlbaum Associates, Hillsdale, NJ.