Tools in Linguistic Theory
Budapest
16 May 2004

Ellipsis Interpretation, Inference and Violable Constraints

Daniel Hardt
CMOL/Computational Linguistics
Copenhagen Business School

1 Introduction

Parallelism Constrains VP Ellipsis Interpretation

(1) John told Sam to feed his$_i$ cat. Harry did too. (told Sam to feed his$_j$ cat)

What are possible referents for his$_i$ and his$_j$?

- Sam - Sam, John - Harry, John - John, Tom - Tom
- *Sam - Harry, *John - Sam, *John - Tom

What Kind of Parallelism?

- **Local Parallelism**

  VPE must be logically identical to antecedent VP(Sag, 1976; Williams, 1977)
  (VPE-specific condition)

- **Non-Local Parallelism**

  Can look at containing constituents to satisfy parallelism(Fiengo and May, 1994; Prüst *et al.*, 1994)
  (General condition: not specific to VPE)

- **Indirect Parallelism**

  - Inference can be used to satisfy parallelism;
  - Need Local Parallelism as well as Non-local Parallelism(Rooth, 1992)

The Simplest Possible Theory

Fox's proposal

- No VPE-specific condition is required: instead, the general Non-local Parallelism condition is combined with general constraints on inference.(Fox, 2000)
- Two constraints on inference
  1. **Trigger:** Inference must be triggered by Accommodation-Seeking Material (ASM)
  2. **Minimality:** Inference must be minimal (minimality defined in terms of changes in lexical items)

The Plot

- **Background**
  - 2. Three Types of Parallelism
  - 3. Fox's Proposal: Constrained Indirect Parallelism
- **4. The New Proposal**
  - Ordered Violable Constraints:
    - AGREE
    - COHERE (Non-local Parallelism)
    - AVOID INFERENCE
    - STAY (Avoid NP-Raising)
  - 5. Eliminating Fox's Trigger Condition
  - 6. Minimal Inference Defined in Terms of Model Simplicity
  - 7. Strict Blocks Non-local Sloppy Because STAY Penalizes Lambda Abstraction
- **8. Additional Topics**
  - VPE Survey: Linguist-GRID.org
  - Fox's Trigger Condition Revisited
  - Is Lambda Abstraction Syntactic Movement, or Inference?

2 Background: Three Types of Parallelism

Local Parallelism (Sag/Williams)

- VPE must be identical to antecedent VP
- A (deleted) VP has a lambda binder

(2) (Sam-Sam) John_1 [λ x.x told Sam_2 to feed his_2 cat]. Harry_3 did too.[λ x.x told Sam_2 to feed his_2 cat]

(3) (John-Harry) John_1 [λ x.x told Sam_2 to feed x's cat]. Harry_3 did too. [λ x.x told Sam_2 to feed x's cat]

Non-parallel readings: John-Sam, Sam-Harry, are ruled out by Local Parallelism.

(4) John_1 said Mary_2 hit him_1. Harry_3 did too

(5) (sloppy ok) John_1 [λ x.x said Mary_2 hit him_1]. Harry_3 did too [λ x.x said Mary_2 hit x]

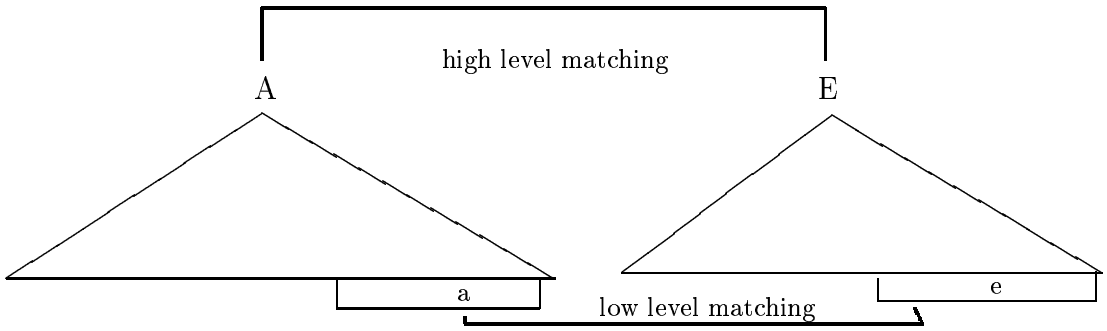
(6) John_1 said Mary_2 hit him_1. Harry_3 said she_2 did too.

(7) (sloppy no good) John_1 said Mary_2 [λ x. x hit him_1]. Harry_3 said she_2 did too. [λ x.x hit him_3]

- With Local Parallelism, (5) permits the sloppy reading, because the VPE is identical to the antecedent VP – [λ x.x said Mary_2 hit x].
- Local Parallelism doesn't permit sloppy for (7) – the VPE must be [λ x.x hit him_3], which is not identical to the antecedent VP.

Non-local Parallelism

- Ellipsis requires two conditions: Local Parallelism, and Non-local Parallelism (Fiengo and May, 1994; Rooth, 1992)



First John told Mary_1 I was [bad-mouthing her_1] then he told SUE_2 I was [bad-mouthing her_2]

Two parts to Rooth's theory of Non-local Parallelism:

- Alternative Sets
- Sloppy Identity and NP-raising (lambda abstraction)

Alternative Sets

- **Rooth 92 – Alternative Sets, applied to ellipsis:**
  Take an ellipsis site e with an ellipsis antecedent a in the discourse. Ellipsis requires that there be some phrase E containing the ellipsis e and some phrase A containing the ellipsis antecedent a such that $[\![A]\!]$ is an element of F(E).
- **Focus semantic value of α, F(α):** The set of denotations produced by substituting all elements of the appropriate semantic type for every focused element in α. (Rooth, 1985)

(8) A = John left.

(9) E = HARRY did too.(leave)

- Focus Value of E
- F(E) = { P | $\exists$ x.P = x did (leave)}

The Alternative Set of propositions of the form *x did leave*. Since "John left" is an element of that set, ellipsis is licensed.

Sloppy Identity and NP-Raising

(10) strict: First John_1 told Mary_2 I was bad-mouthing her_2. Then he told SUE_3 I was. (bad-mouthing her_2)

(11) sloppy: First John_1 told Mary_2 I was bad-mouthing her_2. Then he told SUE_3 I was. (bad-mouthing her_3)

It is clear that Rooth's Alternative Set Parallelism Condition permits the strict reading, since the Focus Value of E is in the set of propositions of the form *John told x I was bad-mouthing Mary.*
To derive the sloppy reading, it is necessary to form a lambda abstract, by raising the NP *SUE*, resulting in the following:

- SUE, λ x.he told x I was bad-mouthing her

Now the pronoun *her* can be bound by the lambda operator, giving

- SUE, λ x.he told x I was bad-mouthing x

with the focus value:

- F(E) { P | $\exists$ y.P = y, λ x.he told x I was bad-mouthing x}

Now A is an element of F(E), as desired.

Indirect Parallelism: Inference

(Rooth, 1992) argues that inference is sometimes required for Non-local Parallelism, because of the following example:

(12) First John told Mary_1 I was bad-mouthing her_1, and then SUE_2 heard I was. (bad-mouthing her_2)

- **Problem:** VPs don't match. Containing clauses don't match either – but sloppy identity is still ok (for some speakers).
- **Solution:** Find a clause A′ that the original A (Antecedent-Clause) entails, so that parallelism succeeds.
- Inference: *if A tells B C, then B hears C.*
- *John told Mary I was bad-mouthing her* ⇒ *Mary heard I was bad-mouthing her.*

Rooth argues that Local Parallelism must still be maintained for ellipsis, because there are differences between ellipsis and deaccenting, as shown by the following examples:

(13) *First John told Mary I was bad-mouthing her, and then Sue did. (heard I was bad-mouthing her)
John told Mary I was bad-mouthing her ⇒ *Mary heard I was bad-mouthing her.*

(14) *First John told Mary I was bad-mouthing her, and then Sue did. (spoke)
John told Mary I was bad-mouthing her ⇒ *John spoke.*

(13) satisfies Rooth's Indirect Parallelism Condition, just like (12) does. But in (13), the elided VP is not identical to its antecedent.
So there are two conditions:

1. Indirect (Non-local) Parallelism (applies generally)
2. Local Parallelism (applies specifically to VPE)

(*Note:* Local Parallelism cannot be retained in the form given by Sag/Williams, since it no longer says when indices can change. So it has to be something like what is proposed by (Fiengo and May, 1994).)

3 Fox's Proposal

Inference is subject to two constraints:

- **trigger:** Inference must be triggered by *accommodation-seeking material* – deaccented overt material in E that is not present in A. (Fox, 2000)[p 99]
- **minimality:** Inference must be *minimal* – inferred A′ must be as close as possible to A, while being parallel to E. (Fox, 2000)[p 98].

The **trigger** condition explains examples (13) and (14) – there is no accommodation-seeking material.
This makes it possible to eliminate Local Parallelism.

4 The New Proposal

- Ordered Violable Constraints:
  - AGREE
    This rules out, e.g., "Martha danced together", while permitting "Martha and Irv danced together"
  - COHERE
    The only aspect of COHERE I will rely on is Non-local Parallelism, as defined above; intuitively, the requirement is that two sentences must share a common background theme (lambda abstract)
  - AVOID INFERENCE
    Violations of this are ordered, based on an ordering of LF's: For example, If reading A involves no inference, reading B involves the inference $\alpha \rightarrow \beta$, and reading C involves the inference $\alpha \rightarrow \gamma$, and β is closer to α that γ is, then A has no violations, B has one and C has two. As explained below, LF's are ordered according to the simplicity of the corresponding models.
  - STAY
    Avoid NP-raising, which is the mechanism underlying formation of lambda abstracts.

5 Eliminating the Trigger Condition

- Rooth (92): Deaccenting allows inference when ellipsis does not. Therefore, Local Parallelism is still required, in addition to Non-local Parallelism
- Fox: need ASM (as described above) to trigger inference
- My Proposal: these effects follow from the fact that COHERE is ranked higher than AVOID-INFERENCE

(15) First John told Mary I was bad-mouthing her, and then

(16) a. Sue did. (heard I was bad-mouthing her)

b. Sue did. (told Mary I was bad-mouthing her)

Inference needed for (16)a:
John told Mary I was bad-mouthing her ⇒ *Mary heard I was bad-mouthing her.*

(For now, we leave the strict reading out of consideration. Below I argue that strict blocks sloppy in examples like this.)

	AGREE	COHERE	AVOID-INFERENCE	STAY
(16)a *Sue heard I was bad-mouthing her*			*	
☞ (16)b *Sue told Mary I was bad-mouthing her*				

(17) John talked to every woman he saw.

(18) a. Bill did too. (talked to every woman he saw)

b. Bill did too. (talked to many women)

	AGREE	COHERE	AVOID-INFERENCE	STAY
☞ (18)a *Bill talked to every woman he saw*				
(18)b *Bill talked to many women*			*	

- The claim that COHERE outranks AVOID-INFERENCE explains the above facts, without resorting to Fox's ASM trigger condition.

Inference without Lexical Trigger

AGREE can also trigger inference.

(19) Irv and Martha wanted to dance together, but Martha couldn't, because her husband was there.(Webber, 1978)

(20) a. Martha couldn't. (dance together)

b. Martha couldn't. (dance with Irv)

Inference needed for (20)b:

(21) Irv and Martha wanted to dance together ⇒ Martha wanted to dance with Irv

- Here there is no ASM. Instead, there is a conflict between the VP antecedent "dance together" and the singular subject Martha: this gives a high-ranked AGREE violation.
- This further supports using ranked constraints instead of explicitly defining a trigger condition for inference – these effects are not predicted by Fox's trigger condition.

	AGREE	COHERE	AVOID-INFERENCE	STAY
(20)a *Martha couldn't (dance together)*	*			
☞ (20)b *Martha couldn't (dance with Irv)*			*	

6 Defining Minimal Inference

Background

- **Preference for Minimal Inference:** Given an Antecedent Clause A, and two LF's, A′ and A″ that are inferable from A, there is a preference to infer the closer of the two.
- **Fox's View:** According to Fox, A′ is closer to A than A″ is, when the accommodated material of A′ "is a proper subset of the accommodated material in A″", where accommodated material is viewed as a set of lexical items (p 98-99). In other words, seek to find the inference that changes the fewest lexical items.
- **Model Simplicity:** A closer inference is one that has a more similar minimal model.

Example: Scope Parallelism and Inference

Consider the following examples:

(22) A doctor saw every patient. A NURSE did too. [saw every patient]

- Scope ambiguity should be resolved in parallel in both sentences. (Sag, 1976)

1. (a doctor x) (every patient y) x saw y. (a NURSE x) (every patient y) x saw y.
2. (every patient y) (a doctor x) x saw y. (every patient y) (a NURSE x) x saw y.
3. *(a doctor x) (every patient y) x saw y. (every patient y) (a NURSE x) x saw y.
4. *(every patient y) (a doctor x) x saw y. (a NURSE x) (every patient y) x saw y.

(23) A doctor saw every patient. A NURSE saw many patients, too.

Same pattern:

1. (a doctor x) (every patient y) x saw y. (a NURSE x) (many patients y) x saw y.
2. (every patient y) (a doctor x) x saw y. (many patients y) (a NURSE x) x saw y.
3. *(a doctor x) (every patient y) x saw y. (many patients y) (a NURSE x) x saw y.
4. *(every patient y) (a doctor x) x saw y. (a NURSE x) (many patients y) x saw y.

Deaccented *saw many patients* triggers inference.

Here are the relevant LF's:

- A: (a doctor x) (every patient y) x saw y
- A': (a doctor x) (many patients y) x saw y
- A'': (many patients y) (a doctor x) x saw y

Reading 1 is correctly licensed by the inference, $A \Rightarrow A'$:

- (a doctor x) (every patient y) x saw y. $\Rightarrow$ (a doctor x) (many patients y) x saw y.

(Note that this inference is not valid if there are no patients, or a very small number of patients. We make a background assumption that this is not the case.)

But, we must block the inference $A \Rightarrow A''$:

- (a doctor x) (every patient y) x saw y. $\Rightarrow$ (every patient y) (a doctor x) x saw y. $\Rightarrow$ (many patients y) (a doctor x) x saw y.

This inference would incorrectly derive Reading 3, with non-parallel scopes.

- Fox's definition, in terms of differences in lexical material, fails to give the right result – for both A$'$ and A$''$, there is a single lexical item which differs, namely "many".

Ordering Inferences by Minimal Models

- **Claim:** A$''$s minimal model is closer to A's minimal model than A$''$s.

First Order Model:

- **Domain** of model M, dom(M): a set of individuals.
- **Interpretation Function** of M, I(M): a function from relation symbols to sets of n-tuples of elements of the Domain. Equivalently, one can treat I as a set of *assertions* about individuals in the domain, as I will do here.

Ordering on models: M1 $\leq$ M2 iff card(dom(M1)) $\leq$ card(dom(M2)) and I(M1) $\subseteq$ I(M2). (Gardent and Webber, 2001; Gardent and Konrad, 2000; Konrad, 2004)

- *M1 is below M2 if M1 has fewer individuals than M2 and the assertions in M1 are a subset of those in M2*

Minimal Models: for any LF α, M ϵ Min-models(α) iff M satisfies α and no model M$'$ that satisfies α is below M in the model ordering.
Ordering on LF's: $\alpha < \beta$ iff $\exists$ m_1 ϵ Min-models(α), m_2 ϵ Min-models(β) $m_1 < m_2$ AND NOT ($\exists$ m_2 ϵ Min-models(β), m_1 ϵ Min-models(α) $m_2 < m_1$).

- *α is less than β iff α has a minimal model less than a minimal model of β, and β does not have a minimal model less than any of α.*

Closeness of LF's:

closer(A,B,C) "B is closer to A than C is" iff B $\leq$ A AND (C $\leq$ B OR (C $\not\leq$ A AND A $\not\leq$ C)).
in words: "B is below A; C is either below B or unrelated to B"

Blocking Inferences: if closer(A,B,C), then $A \Rightarrow B$ will block $A \Rightarrow C$.

Blocking the Undesired Inference

The Models

Background assumptions: there are at least 4 nurses, doctors and patients. "Many" means at least 3.

M. Min model of A: (a doctor x) (every patient y) x saw y

```
d1,d2,d3,d4  doctor(d1), doctor(d2), doctor(d3), doctor(d4)
p1,p2,p3,p4  patient(p1),patient(p2),patient(p3), patient(p4)
n1,n2,n3,n4  nurse(n1),nurse(n2),nurse(n3),nurse(n4)

saw(d1,p1) saw(d1,p2), saw(d1,p3), saw(d1,p4)
```

M′. Min model of A′: (a doctor x) (many patients y) x saw y

```
d1,d2,d3,d4  doctor(d1), doctor(d2), doctor(d3), doctor(d4)
p1,p2,p3,p4  patient(p1),patient(p2),patient(p3), patient(p4)
n1,n2,n3,n4  nurse(n1),nurse(n2),nurse(n3),nurse(n4)

saw(d1,p1) saw(d1,p2), saw(d1,p3)
```

M″. Min model of A″: (many patients y) (a doctor x) x saw y

```
d1,d2,d3,d4  doctor(d1), doctor(d2), doctor(d3), doctor(d4)
p1,p2,p3,p4  patient(p1),patient(p2),patient(p3), patient(p4)
n1,n2,n3,n4  nurse(n1),nurse(n2),nurse(n3),nurse(n4)

saw(d1,p1) saw(d2,p2), saw(d3,p3)
```

- We have seen that M′ is below M, while M″ is not related to M′. This is because M″ has assertions (eg, saw(d2,p2)) that are not in M′, and M′ has assertions (eg, saw(d1,p2)) that are not in M″. So their interpretations are not related by the subset relation, and therefore the models M′ and M″ are unordered.
- This explains why the inference from A to A′ blocks the inference from A to A″, as desired.

7 Strict Blocks Sloppy

Local Parallelism is Still Relevant!

- Despite several decades of counterexamples and counter-proposals, the Sag/Williams Local Parallelism account remains remarkably influential, described, e.g., by (Schwarz, 2000) as "standard in the semantic literature".
- *Linguist-GRID.org* survey: the fully acceptable examples were *exactly* those that were consistent with Local Parallelism.
- **Generalization:** in the absence of Local Parallelism, strict blocks sloppy.
- **Proposed Account:**
  - In general, NP-raising is required to produce a lambda binder
  - NP-raising violates STAY
  - Stipulation: a VPE always has a lambda binder

Strict and Sloppy OK with Local Parallelism

(24) John$_1$ loves his$_1$ cat. Harry$_2$ does too. (λ x.x loves his$_1$ cat)

(25) John$_1$ loves his$_1$ cat. Harry$_2$ does too. (λ x.x loves x's cat)

[Lambda binder is obligatory for elided VP.]

	AGREE	COHERE	AVOID-INFERENCE	STAY
☞ *strict Harry$_2$, (λ x.x loves his$_1$ cat)*				
☞ *sloppy Harry$_2$, (λ x.x loves x's cat)*				

Sloppy Blocked without Local Parallelism

(26) (strict) First JOHN$_1$ told MARY$_2$ I was bad-mouthing her$_2$, and then he$_1$ told JANE$_3$ I was. (bad-mouthing her$_2$)

(27) (sloppy)First JOHN$_1$ told MARY$_2$ I was bad-mouthing her$_2$, and (then) JANE$_3$,λ x. he$_1$ told x I was. (bad-mouthing x)

Strict reading is preferred, because Sloppy requires NP raising, to form lambda-abs [JANE$_3$,λ x. he$_1$ told x I was. (bad-mouthing x)].

	AGREE	COHERE	AVOID-INFERENCE	STAY
☞ (26) *strict:* he$_1$ told JANE$_3$ I was. (bad-mouthing her$_2$)				
(27) *sloppy:* JANE$_3$,λ x. he$_1$ told x I was. (bad-mouthing x)				*

(28) First JOHN told his first year students he would tutor them, and then SUE told JANE I was. (bad-mouthing Mary/Jane)

Strict reading is preferred, because Sloppy requires NP raising, to form lambda-abs [JANE, λ x.SUE told x I was. (bad mouthing x)].

If Strict is Unavailable Sloppy is no Longer Blocked

(29) John told nearly every student$_1$ that I was bad-mouthing her$_1$. But he didn't tell SUSAN$_2$ that I was. (bad-mouthing her$_2$)

(30) I wouldn't expect John to tell ANY student$_1$ that I was bad-mouthing her$_1$. I certainly don't think he'd tell SUSAN$_2$ that I was. (bad-mouthing her$_2$)

- The pronoun *her$_1$* is bound by a quantifier in the first sentence.
- So the strict reading (bad-mouthing her$_1$) is uninterpretable.
- (I ignore E-type readings, in which a pronoun can have an quantified antecedent despite the lack of c-command. Perhaps this reading requires the violation of a constraint more highly ranked than STAY.)

	AGREE	COHERE	AVOID-INFERENCE	STAY
☞ SUSAN, λ x.But he didn't tell x that I was. (bad-mouthing her$_x$)				*

- The survey did not include such cases – my judgment is that the sloppy reading is completely acceptable here. So for (29), there is a violation of STAY, but since the strict reading is eliminated, the sloppy reading is accepted as the only candidate.

8 Additional Topics

Linguist-GRID.org Survey

- Linguist-GRID.org: web-based tool for interactive linguistic surveys
- Survey advertised on LINGUIST List in April 2004
- 28 subjects rated 33 examples of VP ellipsis on a 4 element scale

0	Fully Acceptable	
1	Closer to Acceptable	?
2	Closer to Unacceptable	x
3	Fully Unacceptable	*

Some Observations

- All Locally Parallel examples are Fully Acceptable
- All Non-local Parallelism examples have intermediate status
- All Indirect Parallelism examples are less acceptable than Non-local Parallelism examples
- Absence or presence of ASM *not* systematically associated with acceptability
- *Complete survey data can be found on the web, at www.id.cbs.dk/~dh or www.linguist-GRID.org*

sdev	mean	Judgment (median)	Strict	Sloppy	Local Parallel	Non-local Parallel	Indirect Parallel	ASM
0.26	0.07							
0.50	0.22							
0.69	0.25							
0.64	0.26							
0.65	0.29							
0.76	0.32							
0.81	0.36							
0.67	0.36							
0.89	0.36							
0.78	0.37							
0.82	0.43							
0.98	0.50							
1.01	0.61							
1.06	0.75							
1.19	1.07	?						
1.21	1.54	×						
1.24	1.57	×						
1.23	1.61	×						
1.12	1.81	×						
1.12	1.81	×						
1.03	1.86	×						
1.07	1.93	×						
1.13	1.93	×						
1.05	1.96	×						
1.18	1.96	*						
0.99	2.25	*						

Fox's Trigger Condition Revisited

(31) * First John told MARY$_1$ I was bad-mouthing her$_1$, but then SUE$_2$ DENIED I was. (bad-mouthing her$_2$)

(32) First John told Mary$_1$ I was bad-mouthing her$_1$, and then SUE$_2$ heard I was. (bad-mouthing her$_2$)

- (Fox, 2000)[p 100]: (31) doesn't permit sloppy reading, because that would require inference, in the absence of ASM – (32) is ok on sloppy reading, because deaccented *heard* is ASM.
- This claim receives some support from survey – the median for (31) is * (unacceptable), with the mean 2.25, while the median for (32) is x (mostly unacceptable), with the mean: 1.81.
- But, Fox's explanation is contradicted by the following examples from the survey:

(33) First JOHN told MARY I was bad-mouthing her, and then SUE INSISTED that I was. (bad mouthing Sue)
mean: 1.81 median x

(34) First people told MARY that I would have to tutor her, and then SUE INSISTED that I would. (have to tutor Sue)
mean: 1.96 median x

(33) is as good as (32), but has same structure as (31).

Is Lambda Abstraction Syntactic Movement, or Inference?

- Lambda abstraction could be considered a valid inference, since it does not change meaning
- Could eliminate STAY, and treat lambda abstraction as AVOID-INFERENCE violations instead
- Wouldn't change results on any examples considered here

9 Conclusion

- The simplest possible theory of ellipsis (Fox 2000): the facts follow from general, independently motivated conditions – parallelism and constrained inference.
- **My Reformulation:**
  - Ordered Constraints: AGREE, COHERE, AVOID-INFERENCE, STAY
- **Benefits:**
  - Eliminate Triggering Condition
  - Capture Blocking Effect: Strict blocks Sloppy unless there is Local Parallelism
  - All three previously defined parallelism theories re-emerge in the ordered constraints theory
- Minimal Inference defined in terms of model simplicity

References

[Fiengo and May, 1994] Robert Fiengo and Robert May. *Indices and Identity.* MIT Press, Cambridge, MA, 1994.

[Fox, 2000] Danny Fox. *Economy and Semantic Interpretation.* MIT Press, 2000.

[Gardent and Konrad, 2000] Claire Gardent and Karsten Konrad. Interpreting definites using model generation. *Journal of Logic, Language and Information*, 1(2):193–209, 2000.

[Gardent and Webber, 2001] Claire Gardent and Bonnie Webber. Towards the use of automated reasoning in discourse disambiguation. *Journal of Logic, Language and Information*, 10:487–509, 2001.

[Konrad, 2004] Karsten Konrad. *Model Generation for Natural Language Interpretation and Analysis.* Springer, 2004.

[Prüst *et al.*, 1994] Hub Prüst, Remko Scha, and Martin van den Berg. A discourse perspective on verb phrase anaphora. *Linguistics and Philosophy*, 17(3):261–327, 1994.

[Rooth, 1985] Mats Rooth. *Association with Focus.* PhD thesis, University of Massachusetts-Amherst, 1985.

[Rooth, 1992] Mats Rooth. Ellipsis redundancy and reduction redundancy. In *Proceedings of the Stuttgart Ellipsis Workshop*, Stuttgart, Germany, 1992.

[Sag, 1976] Ivan A. Sag. *Deletion and Logical Form.* PhD thesis, Massachusetts Institute of Technology, 1976. (Published 1980 by Garland Publishing, New York).

[Schwarz, 2000] Bernhard Schwarz. *Topics in Ellipsis.* PhD thesis, University of Massachusetts-Amherst, 2000.

[Webber, 1978] Bonnie Lynn Webber. *A Formal Approach to Discourse Anaphora.* PhD thesis, Harvard University, 1978. (Published 1979 by Garland Publishing, New York).

[Williams, 1977] Edwin Williams. Discourse and logical form. *Linguistic Inquiry*, 8(1):101–139, 1977.