ELLIPSIS RESOLUTION AND INFERENCE

by

DANIEL HARDT

1 Introduction

Since Webber (1978) it has been known that ellipsis interpretation sometimes requires inference. However, the literature on ellipsis has largely ignored this: prominent accounts such as (Sag (1976); Williams (1977); Dalrymple et al. (1991); Fiengo and May (1994) and Hardt (1999)) provide no mechanism for inference in ellipsis resolution. The proposal in Rooth (1992) is an exception to this: Rooth gives several examples in which inference is required in ellipsis interpretation. He argues that ellipsis is subject to two different kinds of licensing conditions: an ellipsis-specific identity condition and a general focus-background condition. Rooth argues that the focus-background condition interacts with inference, while the ellipsis-specific identity condition does not. Rooth does not propose any constraints on inference. This presents an evident problem of overgeneration, since it is possible to come up with some chain of inferences to license virtually any interpretation of an elliptical expression.

In this paper I will build on a suggestion in Fox (2000) that a constrained inference operation better accounts for the ellipsis facts, and furthermore results in a simpler theory, by eliminating the ellipsis-specific identity condition. I will argue that Fox's proposal is correct, but that its two key claims must be modified. Fox's first claim is that inference must be minimal; but he defines minimality in terms of sets of lexical items. I argue that minimality is properly defined in terms of a simplicity ordering on minimal models, building on notions familiar from Artificial Intelligence and Natural Language Processing. Fox's second key claim is that inference must be triggered in a very specific way – namely, by the existence of new, but deaccented lexical material. I show that this second condition can, and indeed must be removed, because inference is sometimes required in the absence of this trigger. This results in an even simpler theory: an acceptable interpretation is one that results from a minimally required inference.

In what follows, I first present Rooth's two-level theory (Rooth (1992)), in which ellipsis is subject both to a general focus-background condition, and an ellipsis-specific identity condition. Rooth notes that the focus-background condition can involve inference. I turn next to Fox's proposal, which places constraints on the interaction with inference and shows that this makes it possible to eliminate the ellipsis-specific identity condition. Next, I present a problem for Fox's lexically defined minimality, and I argue that inference minimality is best defined in terms of ordering on minimal models. Finally, I show that inference need not be triggered by lexical material in the way Fox suggests. Rather, I propose that inference is permitted if it is required to repair violations of other constraints.

2 Rooth: a Two-Level Theory of Ellipsis

2.1 The Need for a High-Level Condition

Rooth proposes that ellipsis involves a general focus-background relation which must be satisfied. Rooth (1992) This condition is not necessarily restricted to the minimal clause containing the ellipsis site.

Rooth credits Fiengo and May (1994) for first proposing such a two-level architecture for VP ellipsis resolution. Fiengo and May propose that there is both a VP-specific identity condition and a high-level condition, where the high-level condition is not restricted to the minimal clause containing the elided VP. Rooth illustrates the need for a high-level condition with example (1).

(1) First John told Mary$_1$ I was bad-mouthing her$_1$, then he told SUE$_2$ I was. (bad-mouthing her$_2$)

Rooth notes that (1) permits a sloppy reading – "John told Sue I was bad-mouthing Sue". In general, a sloppy reading requires parallel controllers. Thus the sloppy reading is incorrectly ruled out if one compares minimal clauses a and e. Instead, it is necessary to compare containing clauses A and E. Then we can see that the requisite parallelism holds: everything except focused element (SUE) matches.

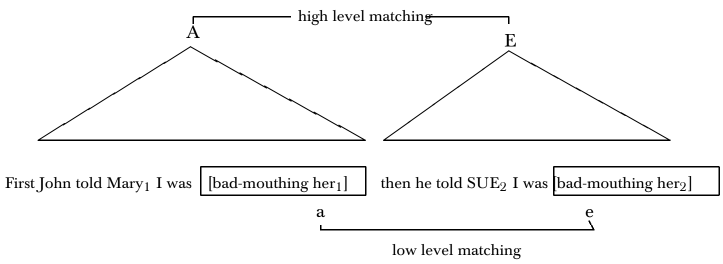

Figure 1: Two-Level Theory of Ellipsis

This is shown in Figure 1.

- **A clause:** John told Mary I was bad-mouthing her
- **Focus Value of E clause:** John told x I was bad-mouthing x

2.2 A Focus-Based High Level Condition

We have seen that the high-level condition is needed to explain the possibility of sloppy identity in the above example. But what exactly are the two conditions? For Fiengo and May, both the low-level and high-level conditions are stated as rather intricate syntactic identity conditions. As we see below, Rooth argues that the high-level condition must be semantic, to account for the fact that it can involve inference.

Rooth gives the following two-level account of ellipsis resolution. An ellipsis-specific identity condition relates the ellipsis clause e to some antecedent clause a. The high-level condition relates some clause E that contains e to some clause A that contains a. More specifically, Rooth's focus-background condition is stated as follows:

Rooth's Focus-Background Condition:
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 or contextually implies a member of F(E).

It is important to note that this Focus-Background Condition applies the same, whether there is ellipsis or deaccenting. It relies on the following technical definition of Focus Semantic value:

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))

2.3 The Need for Inference

Rooth's focus-background condition allows for two possibilities: the antecedent clause must either be an element of the alternative set of E, or *contextually imply* an element. Rooth illustrates the need for inference with examples like the following:

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

Here, the containing clauses do not match: the alternative set for the ellipsis clause is clauses of the form "x heard I was bad-mouthing x", while the antecedent clause is not an element of that set. Rooth argues here that the following inference is required: *if a tells b about c, then b hears about c.* . Thus we have:

A1 *First John told Mary I was bad-mouthing her* $\Rightarrow$
A2 *Mary heard I was bad-mouthing her.*

While *A1* is not an element of the alternative set of the ellipsis clause, *A2* is in the alternative set, and since *A1* entails *A2*, Rooth's condition is satisfied. Without inference, Rooth argues, it isn't possible to account for the sloppy reading.

2.4 The Low-level Condition: Deaccenting vs Ellipsis

Rooth's focus-background condition is different from the condition in theories like (Sag (1976); Williams (1977)), where the elided VP itself must be logically identical to the antecedent VP. Rooth's condition is not specific to VP ellipsis, but holds in general of focus-background structures. Furthermore, even in cases of VPE, it is not restricted to the elided VP itself; it can be applied to any clause containing the elided VP.

Rooth does not offer his focus-background condition as an alternative to the VP identity condition. Rather, he argues that both conditions are required. This position is based on the fact that there is a contrast between deaccenting and ellipsis.

(3) First John told Mary I was bad-mouthing her, and then SUE heard I was bad-mouthing her.

(4) First John told Mary I was bad-mouthing her, and then SUE did. *(heard I was bad-mouthing her)

Example (4) cannot be interpreted the same as (3), despite the fact that this would be consistent with the high-level condition. To capture this, Rooth retains a VP-specific condition on ellipsis, so that "heard I was bad-mouthing her" is not identical to the previous syntactic VP, and therefore cannot be elided.

Rooth has argued convincingly that a high-level condition is required for inference interpretation, and that this condition must interact with inference. His proposal is appealing in that the high-level condition is an independently motivated, general condition on focus-background structures in discourse. There are two aspects to the proposal that are less appealing: first, since it retains a low-level condition, his proposal involves a complication to the theory of ellipsis. Secondly, Rooth's theory is completely unconstrained with respect to inference – any contextually inferrable representation can satisfy the focus-background condition. This casts doubt on the empirical content of the overall theory, since one might imagine that many absurb readings might be licensed by appealing to long chains of inference.

These problems are addressed in Fox's proposal, which we consider next. In this proposal, there are constraints placed on inference, thus addressing one of the above problems. Furthermore, we will see that this in turn makes it possible to eliminate the other problem: we will be able to eliminate the ellipsis-specific identity condition.

3 Fox: Triggered, Minimal Inference

I turn now to the proposal from Fox (2000), for integrating triggered, minimal inference with ellipsis interpretation. Fox takes as his point of departure Rooth's focus-background theory. As we have seen above, this theory relates a clause E (containing ellipsis or deaccenting) to some antecedent clause A. Fox divides this condition into *direct parallelism* and *indirect parallelism.* For direct parallelism, A must be an element of the alternative set of E, which basically means that A must match E, except for any focused elements of E. For indirect parallelism, there can be some clause A′ that satisfies direct parallelism with E, where A′ did not appear in the discourse, but can be inferred from A. Fox then goes on to propose the following two constraints on inference:

(5) **trigger:** Inference must be triggered by *accommodation-seeking material* – deaccented overt material in E that is not present in A. (Fox (2000)[p 99])

(6) **minimality:** Inference must be *minimal* – inferred A′ must be as close as possible to A, while matching E. (Fox (2000)[p 98])

Below, I will examine Fox's definition of *miminal inference*, and argue that it must be replaced. I will also argue that the triggering condition must be modified.

3.1 Eliminating the Ellipsis-specific Condition

The first application of Fox's theory is that it makes it possible to eliminate the ellipsis-specific identity condition in Rooth's theory. Fox argues that the difference between deaccenting and ellipsis follows from constraints on inference. Consider again examples (3) and (4), repeated here:

(7) First John told Mary I was bad-mouthing her, and then SUE heard I was bad-mouthing her.

(8) First John told Mary I was bad-mouthing her, and then SUE did. *(heard I was bad-mouthing her)

Rooth argues that the reading in (8) in ruled out by the ellipsis-specific identity condition: *heard I was bad-mouthing her* is not identical to any overt VP in the discourse. Fox argues that the focus-background condition is violated by that reading, because it requires inference, and inference is not permitted here, since there is no accommodation-seeking material to trigger it, while in (7), the deaccented "heard" triggers inference.

Thus while Rooth appeals to the ellipsis-specific identity condition to capture differences between ellipsis and deaccenting, Fox argues that these differences result from the independently required triggering condition on inference.

3.2 Lexically-based Minimality

For Fox, minimal inference is defined in terms of sets of lexical items. Fox uses the term *accommodation* for inferences, and gives the following definitions:

(9) Accommodation of β_{AC} must be minimal given thea accommodation-seeking material α.

(10) a. An accommodation β_{AC} is minimal given α, if there is no alternative accommodation to β_{AC}, β', such that β' contains α and β' is closer to β_A than β_{AC} is.

b. β' is closer to β_A than β_{AC} is, when the accommodated material of β' is a proper subset of the accommodated material in β_{AC}.

c. the accommodated material of an accommodation β consists of the lexical material that is present in β and absent in β_A

This means that, for each potential inference one determines the set of lexical items that are added by the inference. This is termed the *accommodated material* for the inference. Then inferences are partially ordered by the subset relation, applied to their associated *accommodated material.*

4 A Problem for Lexical Minimality

I will use the following examples to argue that Fox's Lexical Minimality must be changed:

(11) A doctor saw every patient. A NURSE saw every patient, too.

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

As observed by many authors (see Asher et al. (2001) and references therein) scope ambiguity should be resolved in parallel in both examples: if "a doctor" takes wide scope, "a nurse" also takes wide scope; if "a doctor" takes narrow scope, so does "a nurse". For (11), this follows from direct parallelism. But for (12), direct parallelism fails, since "many" does not match "every".

Let us see how Fox's theory applies to this. We focus on the case where "a doctor" takes wide scope. Since "many" is deaccented, it provides accommodation-seeking material, which is a trigger for inference. We can then infer A′ (*A doctor saw many patients*) from A (*A doctor saw every patient*). (Note that this only follows if there is a presupposition that there are many patients.) The problem is that A″, with inverse scoping, can also be inferred from A. But this would then license a non-parallel reading. Here are the three relevant LF's:

A		(A doctor x) (every patient y) x saw y.
A′	Parallel:	(A doctor x) (many patients y) x saw y.
A″	Non-parallel:	(many patients y) (A doctor x) x saw y.

It is necessary to rule out the inference giving rise to A″. But this cannot be done in terms of differences in lexical material, since for both A′ and A″, there is a single lexical item which differs, namely "many". Thus, Fox's Lexical Minimality fails in this case.

5 Model-Based Minimality

I would like to propose an alternative notion of minimal inference, based on ideas about model minimality familiar from AI and NLP (Gardent and Webber (2001); Gardent and Konrad (2000)). A model for a logical formula gives a way of making that formula true. Since a model consists of individuals and atomic facts about those individuals, it is natural to order models in terms of how many individuals and facts they contain. This gives a way of indirectly ordering inferred formulas, in terms of the ordering on their corresponding minimal models.

I will first show how minimal inference is defined in terms of model minimality, and then show how this gives the desired result for our example. Recall that we wanted to permit an inference from A to A, while blocking an inference from A to A″. I will show that, in terms of the model-based ordering, A′ is closer to A than A″ is, which I will write as follows: $closer(A, A', A'')$.

5.1 Model Minimality

Models are ordered in terms of the cardinality of their domains, together with subset relations on the interpretation function. The ordering of LF's is made to reflect the ordering of the corresponding minimal models. I will define a closeness ordering on LF's on the basis of a closeness ordering of the corresponding minimal models.

First, I define the set of minimal models for an LF, as follows. [1]

We begin with an ordinary **First Order Model**, which is defined as a Domain and an Interpretation Function, as follows:

- **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. Here, we treat I as a set of *assertions* about individuals in the domain.

[1] This is adapted from Gardent and Webber (2001)[page 8]. See also Gardent and Konrad (2000); Konrad (2004).

We order models in terms of the cardinality of the Domain and subset inclusion of the Interpretation.

Ordering on models

M1 $\leq$ M2 iff card(dom(M1)) $\leq$ card(dom(M2)) and I(M1) $\subseteq$ I(M2).

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

A minimal model is minimal with respect to the above ordering.

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. Next, we can order LF's in terms of the ordering of their corresponding minimal models.

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.*

Finally, we use the ordering to compare the *closeness* of LF's to one another.

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 $\not\lesssim$ A AND A $\not\lesssim$ C)).

in words: "B is below A; C is either below B or unrelated to B"

This is the basis for blocking inferences to "more distant" LF's.

Blocking Inferences

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

5.2 Minimal Inference in our Example

Now we can return to our example. Note first that I assume that there are at least four doctors, patients, and nurses, and that many means "at least 3". While this may seem arbitrary, I believe some such assumption is required here: the use of "every" presupposes the existence of a domain of some reasonable size. For example, it is a bit odd to say "every student asked a question" if there are just two students; in this case "both" is a more felicitous determiner.

So each minimal model will contain the following information:

d1,d2,d3,d4	doctor(d1), doctor(d2), doctor(d3), doctor(d4)
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)

Where the three models differ is in the *saw* relation:

A	saw(d1,p1), saw(d1,p2), saw(d1,p3), saw(d1,p4)
A′	saw(d1,p1), saw(d1,p2), saw(d1,p3)
A″	saw(d1,p1), saw(d2,p2), saw(d3,p3)

The model of A′ is below A, since the respective interpretations of "saw" stand in a subset relation (and everything else is the same). The model of A″ is not related, since here we don't have a subset relation. In A″, *saw* contains pairs not in A′, and conversely in In A′, *saw* contains pairs not in A″.

This shows that there are models M,M',M'' of LF's A,A',A'' respectively, such that closer(M,M',M''). This satisfies the first clause of the definition of the closeness ordering for LF's: there must exist minimal models in the desired order. The second part of the definition requires that there not exist minimal models with the other ordering. That is, we must show that there are no models M,M',M'' of LF's A,A',A'' respectively, such that closer(M,M'',M'). To see this, note that A' has the following minimal models (I ignore everything except the *saw* relation):

A′

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

In all cases, the *saw* relation must contain three pairs, since we stipulate than *many* means "at least 3". Also, each pair must contain the same doctor, since A' gives "a doctor" wide scope. A'' contains all the minimal models of A'. In addition, it has minimal models in which there is not the same doctor seeing each patient, since for A'' "a doctor" gets narrow scope. There is no minimal model of A'' that is below a minimal model for A' in the model ordering, since each minimal model of A'' must have a *saw* relation with at least 3 elements as well; thus the *saw* relation in the A'' minimal model will either be identical to the *saw* relation in the A' minimal model, or it will be unrelated by the subset relation.

This satisfies the second clause of the definition of the closeness ordering for LF's: there indeed do not exist minimal models with an ordering other than the desired ordering. So we have seen that model minimality gives a semantic basis for ruling out the inference from A to A″; namely, the inference from A to A′ gives rise to a closer minimal model than the inference from A to A″.

6 Generalizing the Inference Trigger Condition

I turn now to Fox's claim that inference must be triggered by accommodation-seeking material, i.e. deaccented overt material in E that is not present in A. In fact, it has long been known that inference is sometimes required in cases where there is no accommodation-seeking material. Consider the following example, from Webber (1978):

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

Here, the antecedent VP is "dance together". This cannot directly provide the meaning of the VPE, which is "dance with Irv". Webber argues that this is made available by the following inference:

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

Here there is no accommodation-seeking material. Instead, there is a conflict between the VP antecedent "dance together" and the singular subject Martha. Presumably this is a kind of semantic number agreement violation. This shows that Fox's trigger condition is too specific. Perhaps the correct statement of the condition might be something like: inference can be performed only if it is required to repair some violation. Fox's trigger condition involves an indication that the focus-background condition will be violated in the absence of inference; but here, we have seen that a semantic number agreement violation also triggers inference.

7 Conclusion

While it has long been known that inference is sometimes required in ellipsis resolution, there has been little work on integrating inference into the ellipsis resolution process. The proposals of Webber (1978) and Rooth (1992) make no attempt to constrain the application of inference. In this paper, I build on the proposal in Fox (2000) to integrate constrained inference with interpretation of ellipsis and deaccenting. This proposal makes it possible to simplify the theory by eliminating the ellipsis-specific identity condition. Fox argues that ellipsis is in fact regulated by a general condition relating focus and background material in discourse – this condition may either apply directly, or it may be mediated by inference. Fox's argument has two crucial elements: inference must be minimal, and it must be triggered. I show that both of these elements must be modified in important ways. While Fox defines minimality in terms of sets of lexical items, I show that a model-based notion of minimality is empirically superior to Fox's lexically-based minimality. Furthermore, I show that Fox's triggering condition must be generalized: inference is not always triggered by explicit lexical material, but rather, is required to repair various types of violations.

References

N. Asher, D. Hardt, and J. Busquets. Discourse parallelism, ellipsis, and ambiguity. *Journal of Semantics*, 18(1), 2001.

M. Dalrymple, S. Shieber, and F. Pereira. Ellipsis and higher-order unification. *Linguistics and Philosophy*, 14(4), August 1991.

R. Fiengo and R. May. *Indices and Identity.* MIT Press, Cambridge, MA, 1994.

D. Fox. *Economy and Semantic Interpretation.* MIT Press, 2000.

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

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

D. Hardt. Dynamic interpretation of verb phrase ellipsis. *Linguistics and Philosophy*, 22 (2):187–221, 1999.

K. Konrad. *Model Generation for Natural Language Interpretation and Analysis.* Springer, 2004.

M. Rooth. *Association with Focus.* PhD thesis, University of Massachusetts-Amherst, 1985.

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

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

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

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

Daniel Hardt
Center for Computational Modeling of Language
Department of Computational Linguistics
Copenhagen Business School
Denmark
Email: dh@id.cbs.dk
Web: http://www.id.cbs.dk/~dh