

The Role of Input in the Acquisition of Generic NPs*

Elisa Sneed
Northwestern University

1. Introduction

Imagine a child and her caregiver looking through a picture book. They turn to an illustration of two monkeys sitting in a jungle tree eating bananas. The child says, “Monkeys.” And the mother says to the child:

- 1) Monkeys eat bananas.

The next day the child and her father are looking through the same book. They come to the page with the monkeys and again the child says, “Monkeys.” The father says to the child:

- 2) The monkeys are eating bananas.

Both the mother and the father have spoken felicitously, but they have made importantly different assertions. The mother has made an assertion about a generic property of monkeys whereas the father has made an assertion about these two particular monkeys. But what does the child understand from her parents’ utterances? Does she notice the morphosyntactic differences? Does she use these cues to guide her interpretation?

The problem of determining an NP’s intended interpretation is compounded by the observation that a given NP form is often compatible with both generic and existential or referential interpretations, as shown in (3) and (4):

- 3) a. The cow has four stomachs.
b. A camel spits three times before sleeping.
c. Passenger pigeons are extinct.
- 4) a. The cow is eating.
b. A camel spit three times.
c. There are passenger pigeons in the natural history museum.

Much recent work by Gelman and colleagues (Gelman 2004; inter alia) has shown that parents use all types of NPs to refer to generic concepts, and has argued that children must therefore rely heavily on contextual cues and world knowledge to determine whether a given utterance is generic. In short, these authors have argued that the input by itself, except in the broadest sense (i.e. children can attend to the use of the plural when their caregiver says, for example, “They have four stomachs” in the presence of a single cow) is uninformative regarding generic interpretation.

“...what we find, through a series of converging studies, is that parental input retains the complexity of mapping, thereby not greatly simplifying the inductive task that children face.”
(Gelman 2004: p. 456)

* I would like to thank Jeff Lidz, Tonia Bleam, Sandra Waxman for their insights and generous contributions to this research. Thanks also to and Sandy Zabell for help with statistical analyses. All errors are mine.

But, in fact, the input is more informative than this suggests. In this paper, I show that a more complete examination of a subset of the transcripts used in these studies (Gelman and Raman 2003; Gelman and Tardif 1998) reveals a systematic distribution of types of noun phrases in the input that a child *could* latch onto to learn about the expression of genericity. We agree with Gelman that parental speech does not make learning easier, but we argue that there are cues to interpretation in the input. Importantly, however, these cues are detectable only if the learner has a rich system of grammatical representation. The systematic patterns exist only at a level of analysis more abstract than surface form. Thus, to the extent that children use these cues in learning about genericity, it follows that they have the relevant abstract representations.

2. Hypothesis and proposal

When we examine the existing research on the acquisition of genericity, we find that some aspects of the grammar of genericity have been overlooked. The current study examines the role of the linguistic environment in the acquisition of the semantics of genericity and in the mapping between form and meaning in this domain by looking at the question from a different perspective, namely focusing on linguistic regularities in the input that have not been fully taken into account, such as NP-form and distribution. I propose a learning algorithm by which a child, who has developed certain grammatical representations, could exploit the input to learn about generics.

I assume, following Chierchia (1995, 1998), that children know the inventory of determiner meanings made available by universal grammar. I also assume that kids know something about the relation between novelty and familiarity on the one hand and indefinite- and definiteness on the other. Roughly, children know that new topics of discourse are often introduced with an Indefinite NP¹ while existing topics of discourse will be referred to with a definite NP and *not* an Indefinite.

Thus, the first step in determining whether or not a given NP is generic involves partitioning the definites from the other NP types. The idea here is that certain determiners will be more likely to be used on the first discourse mention of the noun that they occur with than on the second discourse mention. These will represent the Indefinites because that class of NP is more likely to introduce a discourse referent than to pick up an existing discourse referent. Those determiners that occur equally often on first and second mention are the definites. Once children have been able to divide the NPs in their input between these two groups, they can apply Diesing's Mapping Hypothesis (Diesing 1992) to determine which NPs should get a generic interpretation and which should get an existential interpretation.² In essence, only NPs that occur outside of VP will be interpreted generically.

The Mapping Hypothesis was proposed to account for the available semantic interpretations of Indefinite NPs and to define the role of syntax in determining these interpretations. Diesing uses syntactic structure to determine which NPs in a sentence fall into the restrictive clause and which NPs fall into the nuclear scope in a tripartite representation of quantificational structures

¹ I will use capital-I "Indefinite" to refer to all indefinite NPs including indefinite singulars/plurals and bare plurals. I will use lowercase-i "indefinite" to mean just indefinite singulars/plurals.

² We use Diesing's Mapping Hypothesis for concreteness, but it is important to understand that the precise mechanism of Indefinite interpretation is less important than the basic insight that lies behind this hypothesis, namely, that Indefinites that occur low in the structure are assigned an existential reading whereas Indefinites that occur high in the structure are assigned a generic interpretation. See Rooth 1995, Cohen & Ertshik-Shir 2002 for alternative proposals capturing the same basic insight.

(Kamp 1981, Heim 1982)³. The Mapping Hypothesis (MH) is relevant to our discussion here because it makes a number of predictions that account for the different interpretations available for Indefinite subjects.

The account is based in part on Lewis' observation that Indefinites show quantificational variability (Lewis 1975). In (5a), we interpret the indefinite as referring to most psychologists, whereas in (5b), we interpret the indefinite as referring to few psychologists⁴:

- 5) a. A psychologist usually hates syntax
- b. A psychologist rarely hates syntax.

Kamp (1981) and Heim (1982) argued that these data show that Indefinites are non-quantificational, but rather pick up their quantificational force from other elements in the sentence. More concretely, the Indefinite introduces a variable which is interpreted as the restrictor of the quantificational adverb:

- 6) a. usually [psychologist(x) -> hate (x, syntax)]
- b. rarely [psychologist (x) -> hate (x, syntax)]

When there is no other overt expression of quantificational force, however, Indefinites allow for two interpretations: a generic one and an existential one, as illustrated in (7).

- 7) a. Political activists are passionate.
- b. Political activists are available.

The sentence in (7a) with a bare plural subject is about activists in general, not any specific activists. However, the sentence in (7b) with the very same bare plural subject displays an ambiguity of interpretation. It can either be about some existing activists that are ready for a particular protest, as in (7b') or it could be about activists in general who are always ready for a protest as in (7b'')

- 7) b'. Political activists are available (for the protest).
- b''. Political activists are available (for any protest).

The different interpretations available in (7) can be explained by the MH.

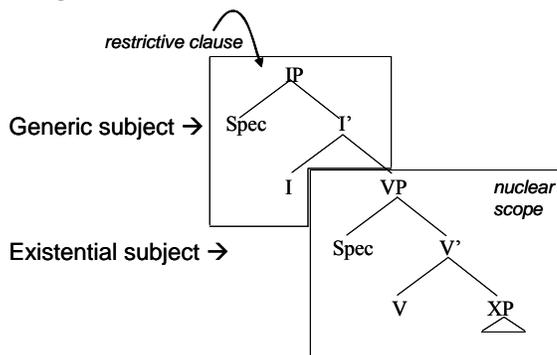
The MH has two components, a semantic component and a syntactic component. In the semantics, there are two quantifiers that can be covertly expressed in a sentence: a generic operator and an existential operator. The syntactic component interacts with the semantic component in the following way. First, a sentence can be partitioned into two domains, the IP and the VP. The IP is the domain of covert generic quantification; the VP is the domain of covert existential quantification. So, material contained in IP but not VP is mapped into the restrictive clause of quantificational adverbs (as in 6) and/or the covert generic operator (as in 7a/b''). Material contained in the VP is mapped into the nuclear scope of such operators (as in 7b'). In

³ The interested reader is directed to Diesing, 1992 for a full discussion of the account.

⁴ Importantly, (a) is not interpreted as meaning "there is a psychologist who usually hates syntax" and (b) is not interpreted as meaning "there is a psychologist who rarely hates syntax," which is what we would expect if the indefinite was interpreted as an existential quantifier.

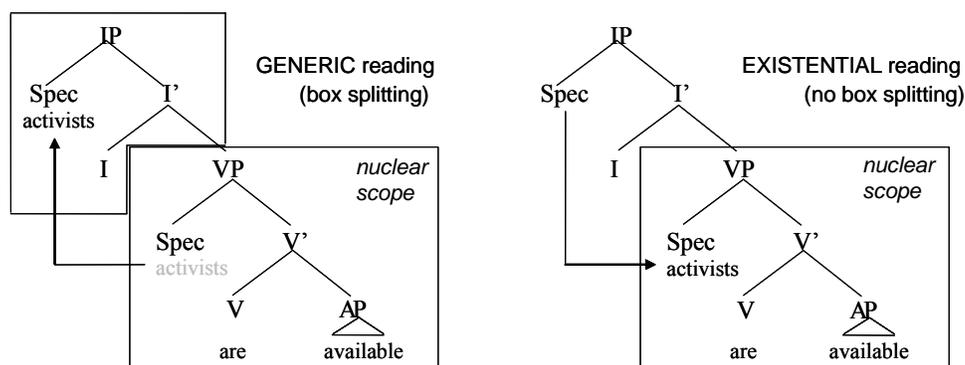
addition, an existential operator can be inserted at the edge of VP to bind any unbound variables such as those introduced by Indefinites. As a result, Indefinite subjects that occur in [Spec,IP] get a generic interpretation in sentences lacking any overt quantification, while those that occur inside VP are interpreted existentially. An important consequence of this system is that objects should always get existential interpretations because they're inside VP⁵.

FIGURE 1 – Box splitting



The difference between the sentence in (7a) and the sentence in (7b), can be explained by the LF position of the subject. The predicate in (7a), *passionate*, is an Individual-level (I-level) predicate, while *available* in (7b) is a Stage-level (S-level) predicate⁶. Only S-level predicates have the VP-internal subject position. So in other words, subjects of I-level predicates will always get a generic interpretation, while subjects of S-level predicates can be interpreted generically or existentially. The two derivations of (7b) are shown in Figure 2.

FIGURE 2 – S-level predicate



The LF difference in English is expressed overtly in languages like German and Dutch that have two overt subject positions (one inside VP and one outside of VP). Note the following

⁵ The only exception to this are objects of psychological predicates (e.g. *like, love, hate, cheerful, nervous, etc.*, denoting psychological states of emotion), which are allowed to adjoin to IP at LF and can thus receive generic interpretations. Again, the reader is directed to Diesing (1992) for the full account.

⁶ I-level predicates correspond, roughly, to predicates denoting permanent properties; S-level predicates correspond, roughly, to predicates denoting temporary properties.

examples from Diesing (1992: 36), where the difference in word order reflects the two available subject positions. The sentence in (8a) provides an example of the subject NP *Linguisten* “linguists” interpreted in the lower, VP-internal position, and (8b) provides an example of the same NP interpreted outside of VP in [Spec,IP].

- 8) a. ...weil ja doch Linguisten Kammermusik spielen.
since PRT PRT linguists chamber music play
“...since there are linguists playing chamber music.”
- b. ...weil Linguisten ja doch Kammermusik spielen.
since linguists PRT PRT chamber music play
“...since (in general) linguists play chamber music.”

Once children have separated definites, to which MH is irrelevant, from Indefinites, they can use the subject-object asymmetry predicted by MH to gain insight into which NPs should be generics (all subjects of I-level predicates, and some subjects of S-level predicates) and which NPs should be existential (some S-level subjects and all objects)⁷. This type of learning algorithm makes the complexity of learning about generics from the input more manageable.

3. Present study

Materials

This experiment was a corpus study of generic NPs in caregiver speech from a linguistic perspective. Susan Gelman provided me with verbatim transcripts from her book-reading context for 20-month-old children⁸.

Procedure

Information about each noun and the utterance it occurred in were entered into a database. Each noun’s database entry included the sentence the noun occurred in, the verb it occurred with and the number of times the noun had been mentioned by a caregiver in a particular transcript. In addition, the type of predicate the noun occurred with, its interpretation in context, its form and its grammatical function were also recorded:

⁷ It is not unreasonable to suggest that kids would be able to profit from the distinction between I-level and S-level predicates in regard to subject interpretation, as children as young as 2-years-old have been shown to be sensitive to the I-level/S-level distinction in their use of the copula (Becker, 2000). At the very least, children should be able to distinguish between subjects (likely to be interpreted generically) and objects (highly likely to be existential).

⁸ Gelman and Tardif (1998) and Gelman and Raman (2003) had also examined toy play and mechanical-toy play, but their results showed that there were significantly higher rates of generic usage in the book-reading context than in either of the other two contexts. For this reason I used only the transcripts from the book-reading context.

FIGURE 3 – Database information

Predicate type	Interpretation	NP-form	Grammatical Function
Stage-level	Generic	Indefinite sg/pl	Subject
Individual-level	Existential	Definite sg/pl	Object
Psychological	Referential	Bare plural	N-predicate
Unknown	Predicate	Other	Other
	Other		

Results

The total number of nouns in the database was 2672. Only 1896 nouns were included in the analyses because 776 token occurred without a predicate⁹. Over 60% of these nouns, 1188 of them, represent a first (767) or second (421) mention.

What we might notice from a first look at the data is that, as shown in Table 1, there are far fewer generics than there are existentials and referentials (or predicates). In fact, there are over four times as many existentials and six times as many referentials as there are generics. What is particularly interesting about this difference, though, is the distribution of the interpretations across different syntactic positions. First of all, referentials are more evenly distributed across subject and object position than are either the generics or the existentials. But generics and existentials also exhibit a different distribution from one another: the former occur as subjects and the latter as objects.

TABLE 1 – Grammatical function x interpretation¹⁰

Reading (%)	Subject (%)	Object (%)	Other (%)
Generic 60 (5)	56 (93)	4 ¹¹ (7)	0 (0)
Existential 291 (24)	1 (0)	165 (57)	126 (43)
Referential 426 (34)	65 (15)	92 (22)	268 (63)
Other 459 (37)	15 (3)	50 (11)	394 (86)

These data do, in fact, present a systematic regularity in the input, but how could a child exploit this pattern to learn about genericity? We now appeal to the learning algorithm discussed in section 2.

The first stage of the learning algorithm requires the child to separate definite NPs from indefinite NPs. This can be done by attending to the form used when the noun is mentioned. Recall that the intuition behind this piece of the algorithm lies in the different functions of these two types of NPs. Only definites can be used anaphorically and so are expected to be used more

⁹ Such examples generally consist of what seems to be labeling. In these examples, the caregiver simply says the noun e.g. “Apple,” or “Apples.” Such tokens were excluded from the analyses because their utterances contained no syntactic information.

¹⁰ A 3x2 χ^2 test of independence with NP-reading (generic vs. existential vs. referential) and position (subject vs. object) as factors shows that the observed distribution is different from the expected distribution ($\chi^2 = 188.76$, $p < 0.0001$).

¹¹ These four tokens of generic objects reflect four direct objects of psych verbs, which, unlike most direct objects, are interpreted generically.

as second mention than are bare plurals and indefinites. Thus, we expect definites to occur relatively equally often as first and second mention, but we expect an asymmetry for Indefinites. This is, in fact, exactly what we see in the input. Table 2, below, shows a difference in distribution for these two classes of NP. Definites occur in relatively equal percentages across first and second mention. Bare plurals and indefinites occur roughly twice as often for the form a noun takes on its first mention as they do for the form it takes on its second mention.

TABLE 2 – Mention x NP-form¹²

NP form (%)	First Mention	Second Mention
Indefinite 664 (54)	455 (69)	209 (31)
Definite 280 (23)	158 (56)	122 (44)
Other 292 (24)	191 (65)	101 (35)

Children can further distinguish the definites from the Indefinites by looking at where in the sentence the noun occurs. Prince (1992) found that subjects are more likely to represent information that has already been mentioned in the discourse, and as a consequence, are more likely to be definite than are non-subjects¹³. Indefinites, on the other hand, canonically introduce new information (which is more natural following given information), and are thus more common later in the sentence. Again, this prediction is borne out in the data. Below in Table 3, we can see that definites occur in roughly equal percentages in subject and object position, while there are nearly four times as many indefinites and bare plurals in object position as there are in subject position.

TABLE 3 – NP-form x grammatical function¹⁴

NP form (%)	Subject	Object	Other
Indefinite 664 (54)	60 (9)	160 (24)	444 (67)
Definite 280 (23)	48 (17)	53 (19)	179 (64)
Other 292 (24)	30 (10)	98 (34)	164 (56)

The second stage of the learning algorithm now can apply. Once a child has separated definites from Indefinites, she knows which nouns are relevant to the Mapping Hypothesis. The definite NPs can be ignored and the distribution of indefinites and bare plurals can be examined. As we saw above, the difference between the sentences in (7a) and (7b) (repeated below in (9)) can be explained by the LF position of the subject. The sentence in (9a) has only one potential subject position: [Spec,IP]. Material contained in IP but not VP is mapped into the restrictive clause of the covert generic operator; thus, the subject of (9a) is interpreted generically. The

¹² A 2x2 χ^2 test of independence with NP-form (indefinite vs. definite) and mention (first vs. second) as factors shows that the observed distribution is different from the expected distribution ($\chi^2 = 12.13$, $p = 0.0005$).

¹³ Prince's analysis considered only "canonical," unmarked subjects (i.e. in English, subjects that occur in a pre-verbal position).

¹⁴ A 2x2 χ^2 test of independence with NP-form (indefinite vs. definite) and position (subject vs. object) as factors shows that the observed distribution is different from the expected distribution ($\chi^2 = 11.83$, $p = 0.0006$).

sentence in (9b) has two available subject positions: [Spec,IP] and [Spec,VP]. If the subject is interpreted inside VP, in the lower subject position, it will be interpreted existentially (as in 9b'). However, if the subject is interpreted outside of VP, it will instead be interpreted generically (as in (9b'')).

- 9) a. Political activists are passionate.
 b. Political activists are available.
 b'. Political activists are available (for the protest).
 b''. Political activists are available (for any protest).

As shown in Table 4, there is a striking difference in where generics and existentials occur in the input. In fact, the input reflects the MH, even more strongly than the MH predicts it should. Nearly without exception, subjects are interpreted outside of VP, as generics, and objects are interpreted within VP, as existentials¹⁵. This generalization is true for both indefinites and bare plurals. What is surprising is the absence of existential subjects, which are predicted to be possible with S-level predicates. In other words, children are not getting evidence for the availability of interpretations like (9b'), where the subject is interpreted inside VP.

TABLE 4 – Form x position predicts interpretation¹⁶

Reading / Position		Generic	Existential
Indefinite			
Subject	45 (8)	42 (93)	1 (2)
Object	120 (21)	0 (0)	92 (77)
Other	396 (71)	0 (0)	87 (22)
Bare plural			
Subject	14 (14)	14 (100)	0 (0)
Object	40 (39)	3 (8)	37 (93)
Other	49 (48)	0 (0)	20 (41)

4. Conclusions

These findings suggest that the input is informative in determining whether an NP is generic or existential (or referential). The results of this corpus study do not reveal anything about how children interpret NPs in utterances that are intended to be generic versus those in utterances that are intended to be existential. However, they do show that there are statistical and distributional cues in parental input that can aid the child in acquisition of generic and referring language. Children get more out of the input about the generic/non-generic distinction than just “morphosyntactic cues, contextual cues and world knowledge” (Gelman 2004, p. 449). These

¹⁵ The three examples of generic objects are with psychological predicates, a finding that is expected based on the MH. Recall that in Table 1, there were four generic objects (all occurring with psychological predicates). The additional token of a generic object, not included in Table 4, which looks only at indefinites, bare plurals and definites was a bare mass noun (*spaghetti*).

¹⁶ A 3x2 χ^2 test of independence with NP-form/ Position (indefinite subject vs. indefinite object vs. indefinite other) and interpretation (generic vs. existential) as factors shows that the observed distribution is different from the expected distribution ($\chi^2 = 215.63$, $p < 0.0001$).

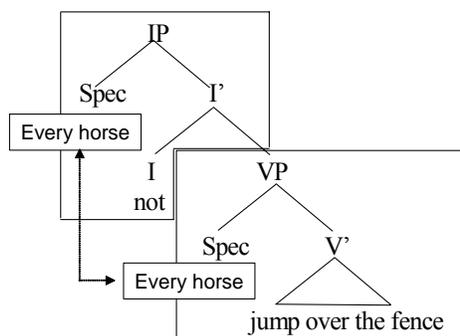
cues hold over abstract categories like grammatical function, predicate type and NP form. Thus, apprehension of these statistical cues requires the child to have made certain abstract grammatical generalizations. Such a statistical analysis over rich grammatical representations could play an important role in the acquisition of genericity.

These findings may also bear on another puzzle in acquisition because they suggest a possible explanation for subject isomorphism in quantification. In sentences with ambiguous quantificational structures, children have been shown to prefer the interpretation conveyed by the surface word order. In other words, children fail to use reconstruction to arrive at the (b) interpretation in sentences like the following, even when this interpretation is true (Musolino, Crain and Thornton, 2001):

- 10) Every horse didn't jump over the fence.
a. For every horse, it's not the case that it jumped over the fence.
b. It's not the case that every one of the horses jumped over the fence.

Recall that while subjects of S-level predicates should be interpretable in [Spec,IP] (where they get a generic interpretation) or in [Spec,VP] (where they get an existential interpretation), children get no evidence for the latter possibility. It may be that the lack of existential subjects masks the availability of the lower subject position. Thus children always interpret the subject in sentences like (10) as taking scope over negation.

FIGURE 5 – Subject isomorphism



5. Future directions

The results of this study offer a piece of evidence for how children derive generic NPs from context, but they open up several other questions: 1) Are these distributions part of child-directed speech or are they a general property of English? 2) Do kids pay attention to these differences? 3) Do they use these in learning? A number of follow-up experiments are underway to address these.

References

- Becker, Misha. 2000. "The Acquisition of the English Copula," in S. Catherine Howell et al. (eds.) *Proceedings of the 24th Annual Boston University Conference on Language Development*. Sommerville, MA: Cascadilla Press, pp. 104-15.
- Chierchia, Gennaro. 1995. "Syntactic Bootstrapping and the Acquisition of Noun Meanings: the Mass/Count Issue," in B. Lust (ed.) *Syntactic Theory and Language Acquisition*. Princeton, NJ: Erlbaum Associates, pp. 301-18.
- Chierchia, Gennaro. 1998. Reference to Kinds across Languages. *Natural Language Semantics*, **6**, 339-405.
- Cohen, Ariel. 2001b. On the Generic Use of Indefinite Singulars. *Journal of Semantics*, **18**, 183-209.
- Cohen, Ariel and Nomi Erteshik-Shir. 2002. Topic, Focus and the Interpretation of Bare Plurals. *Natural Language Semantics*, **10**, 125-65.
- Diesing, Molly. 1992. *Indefinites*. Cambridge, MA: MIT Press.
- Gelman, Susan, John Coley, Karl Rosengren, Erin Hartman, Athina Pappas. 1998. Beyond Labeling: The Role of Maternal Input in the Acquisition of Richly Structured Categories. *Monographs of the Society for Research in Child Development*, **63** (serial no. 253).
- Gelman, Susan and Twila Tardif. 1998. Generic Noun Phrases in English and Mandarin: An Examination of Child-directed Speech. *Cognition*, **66**, 215-48.
- Gelman, Susan and Lakshmi Raman. 2003. Preschool Children Use Linguistic Form Class and Pragmatic Cues to Interpret Generics. *Child Development*, **74**, 308-25.
- Gelman, Susan. 2004. "Learning Words for Kinds," in D. Hall and S. Waxman (eds.) *Weaving a Lexicon*, Cambridge, MA: MIT Press, pp. 445-84.
- Heim, Irene. 1982. *The Semantics of Definite and Indefinite Noun Phrases*, PhD dissertation, University of Massachusetts, Amherst.
- Hollander, Michelle, Susan Gelman and Jon Star. 2002. Children's Interpretation of Generic Noun Phrases. *Developmental Psychology*, **38**, 883-94.
- Kamp, Hans. 1981. "A Theory of Truth and Semantic Representation," in J. Groenendijk, T. Janssen, and M. Stokhof (eds.) *Formal Methods in the Study of Language: Proceedings of the Third Amsterdam Colloquium*. Amsterdam: Mathematical Centre, pp. 277-322.
- Lewis, David. 1975. "Adverbs of Quantification," in E. Keenan (ed.) *Formal Semantics of Natural Language*. Cambridge: Cambridge University Press, pp. 3-15.
- Musolino, Julien, Stephen Crain, Rosalind Thornton. 2001. Navigating negative semantic space. *Linguistics*, **38**, 1-32.
- Prince, Ellen. 1992. "The ZPG Letter: Subjects, Definiteness, and Information-status," in S. Thompson and W. Mann (eds.) *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text*. Philadelphia/Amsterdam: John Benjamins, pp. 295-325.
- Rooth, Mats. 1995. "Indefinites, Adverbs of Quantification and Focus Semantics," in G. Carlson and F.J. Pelletier (eds.) *The Generic Book*. Chicago: The University of Chicago Press, pp. 265-99.