

## Bengali emphatic clitics in the lexicon-syntax interface\*

Josef Bayer and Aditi Lahiri

### 1. Introduction

In this paper, we advocate the view that to provide an adequate account of certain clitic constructions, one must refer to more than one component of the grammatical system. We will argue that the emphatic clitics in Bengali must be licensed by both the morphophonemics of the lexicon and the syntax of logical form, where the latter largely obeys the constraints on overt syntactic movements. After presenting some of the core facts, we will first discuss the lexical account and then turn to the complementary logical-form account. In the last section, we will present our solution to a paradox that appears to arise when both the morphophonemic and semantic aspects of these clitics are considered together.

### 2. The core facts

/o/ and /i/ are the so-called emphatic clitics in Bengali which mean something like 'too' and [+emphatic] respectively. Since only /o/ introduces a new lexical meaning, we will mostly use /o/ for the examples, but /i/ behaves alike in all important respects. At a first glance, /o/ and /i/ seem to adjoin as enclitics to an element of type  $X^o$  which is then the focus of the clitic, e.g.,

---

\* We wish to thank Probal Dasgupta for his advice and Jogamaya Bayer for discussing the Bengali data presented here. We are also grateful to Wim van der Wurff and an anonymous reviewer for their suggestions. Realizing that proposals independently made in Dasgupta (1984, 1987, in press) turned out to be similar in spirit to ours gave us encouragement to pursue this work.

- (1a) *babul- o kha- be*  
 Babul- too eat- [fut, 3rd pers]  
 'Also BABUL will eat'
- b) *babul kha- be- o*  
 Babul eat- [fut, 3rd pers]- too  
 'Babul will also EAT'

In (1a) we see that /o/ can adjoin to a syntactic word and (1b) shows that it can attach to an inflected verb. The clitic can also attach to an overtly inflected N.

- (2) *babul chele- ke- o mereche*  
 Babul boy- [obj]- too beaten-has  
 'Babul has beaten also his SON'

The facts concerning the attachment of the clitic to the verb are more complicated. As we have seen in (1), /o/ can come after the inflected verb, and in most instances, it cannot be added between the stem and the ending. Compare the following pairs of sentences where the second member is ruled out.<sup>1</sup>

- (3a) *mar- i- o*  
 beat- [1pers]- too  
 '(I/we) also BEAT'
- b) \**mar-o-i*
- (4a) *mar- ch- i- o*  
 beat- [prog]- [1pers]- too  
 '(I/we) am/are also BEATING'
- b) \**mar-o-ch-i*

Now consider instances where the clitic can be added before the inflectional ending is attached. Examples are given in (5b) and (5d).<sup>2</sup>

- (5a) *babul cheleke mere- che-o*  
 Babul boy[obj] beaten- has-too  
 'Babul has also BEATEN the boy'
- b) *babul cheleke mere-o-che*
- c) *mer- e- ch- i- l- am- o*  
 beat- [prt]- [prog]- [link]- [past]- [1pers]- too  
 '(I) have also BEATEN'
- d) *mer-e-o-ch-i-l-am*

Although the clitic can be attached between the stem and the inflection, it cannot be inserted between affixes. The choice is binary — either the clitic comes right after the stem, or it must come after all the affixes are added. Thus the forms in (6) are ungrammatical in Bengali.

- (6a) \**mer-e-ch-i-o-l-am*  
 b) \**mer-e-ch-i-l-o-am*

It must be noted that there are no instances of categories other than verbs in which the emphatic clitic could be "infixd". For instance, (7) is ungrammatical. /o/ appears here between a noun stem and the case-marker *-ke*, which is arguably an inflectional ending.

- (7) \**babul chele-o-ke mereche*  
 'Babul has beaten also the BOY'

This restriction also holds for compound-like word formations. Bengali has a verbal noun, which is derived by attaching the suffix *-a* to a V-stem, e.g., *pɔr* 'read' + *a* → *pɔra* 'reading'. Similar to German infinitives such as *rad+fahren* 'bike riding', the verbal noun can incorporate an N°-object into the verb stem involved, e.g., *gɔlpo pɔr-a* 'story reading'. As (8a) below shows, /o/ can adjoin to the verbal noun and select its focus inside, but as shown in (8b), it cannot adjoin to the focused N incorporated. For these examples, imagine a preceding discourse in which someone states that (s)he liked somebody's reading of stories very much.

- (8a) *tader [kobita pɔr- a]- o bhalo laglo*  
 their poetry read-ing- too pleased-has  
 '(I) was pleased by also their reading of POETRY'
- b) \**tader [kobita-o pɔr-a] bhalo laglo*

This also holds true for the so-called dvandva constructions such as in (9) below. The contrasting sentences with different clitic placements are given in (10).

- (9a) *bap ma*  
 father mother  
 'parents'
- b) *uttor dokkhin*  
 north south
- (10a) [*bap ma*]-o  
 b) \**[bap-o ma]*  
 c) [*uttor dokkhin*]-o  
 d) \**[uttor-o dokkhin]*

The interpretation of dvandvas such as (10a) suggests that /o/ attaches to the whole dvandva as indicated by the bracketing, not to its second constituent. Thus (10a) can only mean ‘the parents too’, but not ‘father and also mother’.

Given these observations there seem to be two obvious questions that come to mind. First, what is the domain to which the clitics can attach? Second, if this domain is not identical to the stem to which all affixes can attach, how do the clitics differ from regular affixes?

### 3. Emphatic clitics in the lexicon

The fact that (3b) and (4b) are bad but not (5b) can be accounted for by examining the minimal phonological unit that /o/ can attach to. Observe that /o/ can attach to *mere* but not to *mar*. We will argue that the host of the clitics must minimally be a phonological word. Under this view, *mere* constitutes a phonological word while *mar* is merely a stem. In the following discussion, we will focus on the facts which constitute evidence for differentiating phonological words from stems, indicating also how the clitics themselves are different from regular affixes.

**3.1.** Bengali has a rule which degeminates syllable-initial geminates. The underlying form of the progressive affix is /cch/, a geminate affricate. After a vowel-final stem the geminate is retained (since it can close the preceding syllable), but is degeminated when preceded by a consonant-final stem.

- (11a) *kha- cchi* → *khac.chi* (closure of preceding syllable)  
eat- [prog, 1pers]  
‘(I) am eating’
- b) *mar- cchi* → *mar.chi* (C deleted)

The same holds true after the causative affix /a/ is added to the stem; the geminate is retained as after a stem final vowel.

- (12) *mar- a- cchi* → *ma.rac.chi*  
beat- [caus]- [prog, 1pers]  
‘(I) am having (him) beaten’

Resyllabification is, however, blocked outside the domain of a phonological word; after *mere*, the geminate remains syllable initial and is degeminated.<sup>3</sup>

- (13) *mar- e- cchi* → *me.re.chi* (not: *\*me.rec.chi*)  
beat- [past prt]- [prog, 1pers]

Implicit in this view is the claim that affixes can be added to stems as well as to larger units like words. The compound-like constructions (cf. 9–10) also have affixes added at the end ([*bap ma*]<sub>r</sub> ‘of parents’, [*\*bap-er ma*]). Moreover, the prosodic unit after the addition of a clitic to a phonological word, is still a word to which an affix can be added.

**3.2.** Evidence that *mere* is indeed a phonological word and that there is less cohesion between word+affix than stem+affix comes from reduplicating echo words. In Bengali, an echo word can be formed by reduplicating the entire word except for the initial consonant which is usually replaced by a coronal. The echo word could be interpreted as ‘X and so forth’ with perhaps a slight pejorative tinge. There is a constraint, however, in what can be reduplicated. All stem+affix constructions can be reduplicated, but no stems alone.

(14a) Nominal forms:

<i>pa</i> – <i>ʈa</i>	‘leg’
<i>baʃi</i> – <i>ʈaʃi</i>	‘house’
<i>chele</i> – <i>ʈele</i>	‘boy’
<i>kobita p̄ora</i> – <i>ʈobita p̄ora</i>	‘poetry reading’

b) Verb forms:

<i>mare</i> – <i>ʈare</i>	‘[3rd pers] beat’
<i>khae</i> – <i>ʈae</i>	‘[3rd pers] eat’
<i>mere</i> – <i>ʈere</i>	‘[past part] beat’

c) Inadmissible verb stems:

<i>*kha</i> – <i>ʈa</i>	‘eat’
<i>*mar</i> – <i>ʈar</i>	‘beat’

The affixation pattern also shows the distinction between the stem and the word. The past-participial form *mere* can have the affix *-che* added to it and then become reduplicated; but the *-che* can also be added after reduplication has taken place. Compare the forms under (15).

- (15a) *mereche tereche* '[past part+ 3rd pers] beat'  
 b) [*mere tere*]che  
 c) *marche tarche* '[3rd pers pres] beat'  
 d) \*[*mar tar*]che

This gives more evidence that *mere* is a unit which can stand on its own.

3.3. Evidence that the clitics differ from superficially similar affixes can be obtained from phonological rules of deletion and shortening. Derivative vocalic suffixes can trigger vowel shortening in stem vowels and deletion of vowels; clitics, however, do not trigger such processes. The examples in (16) are taken from Dasgupta (1984).

- (16a) *na:k* 'nose'  
 b) *naki* 'nasal'  
 c) *ra:g* 'anger'  
 d) *ragi* 'angry'  
 e) *na:k-o* 'the nose too' (not: \**nak-o*)  
 f) *ra:g-o* 'anger too' (not: \**rag-o*)  
 g) *no:t* '(male) dancer'  
 h) *no:ti* '(female) dancer'  
 j) *no:t-i* '(male) dancer/[+emphatic]'  
 k) *pagol* 'idiot'  
 l) *pagli* 'mad woman'  
 m) *pogol-i* 'idiot/[+emphatic]'

To summarize, we have shown that the clitic /o/ and /i/ are different from derivational suffixes and have as their minimal host category the phonological word. The latter fact, however, still does not account for the fact that sentences such as (7a) are ungrammatical. It seems unreasonable to suggest that *chele* does not constitute a minimal phonological word.<sup>4</sup> In the next section, we will therefore explore a completely independent line of reasoning.

#### 4. Emphatic clitics in syntax and logical form

Let us make the assumption that the clitics under consideration impose quantificational properties on their morphological/syntactic domain, similar to *only* and *even* in English. As Rooth (1985) and others before him

have argued, a phrase narrowly focused by *only*, *even*, etc., must be interpreted with respect to a quantificational domain. Such a domain is naturally provided by the verb, although not necessarily by the verb. Let us assume that (17a) below is an S-structure, (17b) is the logical form derived from it, and (17c) is a rough semantic representation which transduces the logical form into a proposition with a universal quantifier having scope over it.

- (17a) *we saw only John*  
 b) [*only John*]<sub>i</sub> [*we saw x<sub>i</sub>*]  
 c) For all *x* [*we saw x → x = John*]

*Only John* like *John-o* is a quantifier which must be assigned scope over the clause (proposition) at the level of logical form. Scope assignment, however, is constrained in language-specific ways. For instance, *even* in English cannot appear in an unconstrained fashion, although there is no prima-facie semantic reason which could prevent this.

- (18a) *They have killed* [<sub>NP</sub> *even* [<sub>NP</sub> *my dog*]]  
 b) \**They have killed* [<sub>NP</sub> *my* [<sub>N'</sub> *even* [<sub>N'</sub> *dog*]]]

As Bayer (1988, 1990) has shown, *only*, *even*, etc., and their respective correspondents exhibit different island effects in Dutch, English, and German. (See also Longobardi (in press) for Italian.) Once they are adjoined to an XP which does not correspond to the predicate of a root sentence, this XP must be canonically governed by a verb, or it must connect to a "dynasty" of uniformly oriented governors in the sense of Koster (1986). According to Koster, with the exception of VP, all maximal projections XP of lexical categories are virtual bounding nodes. A bounding domain can, however, be extended when XP is governed by an element which conforms to the basic orientation of government in the language. The direction that counts as basic in a VO-language such as English is →, while it is ← in OV-languages such as German and Bengali. Some examples contrasting English and German will show what is at stake.

- (19a) *John would* [*even* [*talk to MARY*]] (adjoined to VP)  
 b) *John would talk* [<sub>to</sub> [*even* [*MARY*]]]  
 → →  
 (adjoined to NP inside PP)

- (20a) *weil Hans* [*sogar* [*mit MARIA*]] *sprechen würde*  
 → ←  
 (adjoined to PP)

- b) \**weil Hans [mit [sogar [MARIA]]] sprechen würde*  
 ↗ ↖

(adjoined to NP inside PP)

Adopting the rule of quantifier raising (QR) as suggested in May (1977), English allows for *even Mary* to undergo QR because P governs in the same direction as V: the PP ceases to be a bounding node, and (19b) is well-formed. English employs uniformly rightward-looking governors, but German employs a mixed system: Ps that govern to the right, but Vs that govern to the left. Since *sogar Maria* is governed by a rightward-looking governor, the PP counts as a bounding node, and QR is impossible as (20b) shows. The derivation of well-formed logical forms seems to be constrained by the following principles.

- (21a) Focus-sensitive quantifiers (*only, even; nur, sogar; -o* etc.) must have access to a domain of quantification.  
 b) Raising to S (or at least to predicate-level) provides a domain for quantification.  
 c) Governed quantifiers must be (canonically) governed in a dynasty ( $g_1, \dots, g_n$ ) of uniformly oriented governors, up to a tree height where a quantification domain is found.

4.1. Bengali does not have a mixed system as German does. It has exclusively postpositions. Since it is an OV-language, PPs should not lead to island effects. At least partially, this expectation is borne out.

- (22a) *ami sipra- r- o fonge kotha bolbo*  
 I Sipra- [gen]- too with talk say [fut]  
 'I will talk with also SIPRA'  
 b) *ram- er- i fommondhe alocona hocchilo*  
 Ram- [gen]- [emph] about discussion was-taking-place  
 'Discussion took place about RAM'

There are other postpositions, however, which do not allow similar constructions. *Diye* 'with', derived from the past participle of *deowa* 'give'; for example, does not. It is possible that by careful classification an independent difference between Bengali postpositions could be derived which could account for the asymmetry.<sup>5</sup>

4.2. Like in German, Bengali adjectives take NP-complements to the left. As expected, island effects are absent here, too.

- (23) *weil das Gericht [[[dem Patienten] sogar] bekömmlich] ist*  
 since the meal the patient even suitable is  
 'since the meal is even suitable for the PATIENT'

- (24a) *ram- er- o prio*  
 Ram- [gen]- too dear  
 'dear to Ram, too'  
 b) *babul- er- o bhokto*  
 Babul- [gen]- too fond  
 'fond of Babul, too'

4.3. Quantified NPs which are adjoined to  $X^{\max}$  can freely undergo quantifier raising because they are ungoverned. The  $X^{\max}$  to which they belong does not count as a bounding node. The Bengali possessor-NP appears to be adjoined, i.e., not in [SPEC, NP]-position as is the case in English or Standard German. For lack of space we have to simplify somewhat and propose the following structures.<sup>6</sup>

- (25a)  $[_{NP} [_{NP} \textit{amar baba- r}] [_{NP_i} \textit{oi bari}]]$   
 my father [gen] this house  
 'this house of my father's'  
 b)  $[_{NP_i} [_{NP} \textit{my father's}] [_{NP_i} \textit{house}]]$   
 c) \**my father's this house*

Both the grammar of syntactic movement and the grammar of scope assignment behave accordingly, as we will show next.<sup>7</sup>

- (26) NP-split in the syntax  
 a) *tumi [kon lok- ta- r bari] dekhecho?*  
 you which man- [def]- [gen] house seen-have  
 'Which man's house did you see?'  
 b) *[kon lok- ta- r] tumi [bari] dekhecho?*

This is not possible in English. It would violate what has since Ross (1967) become known as the "left branch condition".<sup>8</sup>

- (27a) *[Whose house] did you see?*  
 b) \**[Whose] did you see [house]?*

Exactly the same constraint seems to be at work in logical form, as indicated by the grammaticality difference between (28) and (29):<sup>9</sup>

- (28) NP-split in logical form  
 [amar BABA- r- o] bari bikri hoe gæche  
 my[gen] father- [gen]- too house sold become has-gone  
 ‘Also my FATHER’s house was sold’
- (29a) \*[my FATHER even’s] house was sold  
 b) \*[my father TOO’s] house was sold

Let us assume that N does not lexically govern a structural position such as [Spec, NP]. Despite the canonical direction of government that holds in a language, this guarantees that NP is an island for the specifier, while it is not an island for an adjoined position such as the possessor-NP in (25a).

We can now make the following generalization:<sup>10</sup> only those quantified elements X can undergo quantifier raising which are either adjoined to some YP or linked in a dynasty. A dynasty is built by a chain of successive governors when they govern in the same canonical direction. [<sub>X</sub> X particle] or [<sub>X</sub> particle X] is c(anonically)-governed, if the minimal maximal category dominating X contains a governor G which precedes X in a right-branching language and follows X in a left-branching language. In the following examples, these requirements are not met:

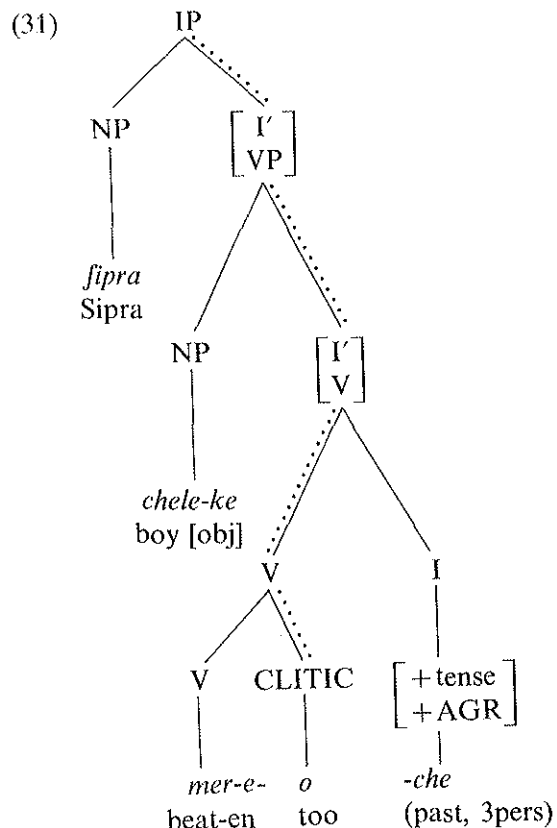
- (30a) \*ami [<sub>NP</sub> nil- o bari] dekhechilam  
 I blue- too house seen-have  
 ‘I have seen a house which is also BLUE’
- b) \*They have killed [<sub>NP</sub> my [<sub>N</sub> even [<sub>N</sub> dog]]]
- c) \*sipra [<sub>NP</sub> chele- o- ke] mereche  
 Sipra boy- too- [acc] beaten-has  
 ‘Sipra has beaten also the BOY’
- d) \*tader [<sub>NP</sub>[<sub>N</sub>[<sub>VP</sub>[<sub>N</sub>kobita- o] [<sub>V</sub> pɔr]]- a]] bhalo laglo  
 their poetry- too read- ing pleased-has

(30a–c) are ruled out because the NP is the minimal maximal category dominating the quantified X, and X fails to be c-governed in NP. The category dominating the quantified *nil-o* in (30a) is the NP *nil-o bari*; analogously, the relevant category in (30b) and (30c) is the NP *my even dog* and *chele-o-ke* respectively. Notice that /o/ in (30c) c-commands only the uninflected noun. After the attachment of the case-suffix *-ke*, which we do not consider to be a governor, the quantified element can only be

the bare noun. The dominating minimal maximal category is the NP across which the quantificational domain cannot be extended. Under the assumption that *kobita-o pɔr-a* in (30d) (= 8b) is a VP which undergoes nominalization due to the affixation of *-a*, *pɔr* c-governs *kobita-o*, but the dominating and intervening NP blocks the formation of a dynasty with the verb (*bhala*)*laglo*. Obviously, the nominalized VP itself does not count as an appropriate quantification domain. Thus, all the examples in (30) arguably constitute “interpretive islands” in the sense of Bayer (1990).

## 5. An apparent paradox

The results achieved in Section 3 left us with an apparent paradox: when the minimal lexical category for the affixation of an emphatic clitic is the phonological word, why is the clitic only allowed inside inflected verbs, but not – as we have just seen – inside inflected nouns? In the light of our discussion in Section 4, the answer to the problem is quite obvious. Following Chomsky (1986), we assume that S is a maximal projection of the inflectional element I(NFL), i.e., S = IP. Apart from this, however, we do not assume that in Bengali I heads a separate node in the syntax. Rather, I being an inflectional feature seems to merge with V such that it becomes the formal head of IP, while V becomes the semantic head.<sup>11</sup> Wherever the emphatic clitic attaches to the inflected verb – observing the morphophonemic constraint – it will quantify over an element that is the head of the clause. Irrespective of the availability of a c-governor, the cliticized/quantified past participle or the cliticized/quantified [V + I] will automatically percolate to a domain where quantification can take place. (31), which means ‘Sipra has also BEATEN the boy’, symbolizes this percolation process with a dotted line.



### Notes

1. *mar* is the V-stem.
2. *mer-e* is the form of the past participle.
3. The stem vowel of verbs alternates in specific morphological environments.
4. One piece of evidence is that the above discussed reduplication rule produces better results when the uninflected noun *chele* is affected: [*chele tele-ke* versus ??[*chele-ke tele-ke*].
5. See Kayne (1981) where the question is raised why English but not French or other Romance languages allow for preposition stranding. Kayne argues that prepositions in English are structural governors, but not in French. Bengali could be a language with a "mixed system" in the sense of Kayne's proposal. The details of such an account, however, would have to be worked out.
6. A more refined analysis could probably be given with the hypothesis that our NP is in fact a determiner phrase (DP). See Abney (1987).
7. See also the similar case of Hungarian (Szabolcsi, 1984). Notice that Bengali keeps WH-words in situ like Chinese and Japanese. What looks like WH-movement in (26b) is rather a case of "scrambling".

8. In the framework of Chomsky (1986) one could account for the left-branch condition because adjunction to arguments (here NP) for purposes of movement is generally not allowed in this theory.
9. Notice that post-NP *even* as well as *too* is generally acceptable, as in *my FATHER even would agree on that* and *my father TOO would agree on that*. Under the assumption that the English genitive *-s* is a clitic which is unselective with respect to its host (cf. Zwicky—Pullum 1983), the examples in (27) cannot easily be excluded otherwise.
10. The predicate "quantified" here pertains only to the cases under consideration, namely focusing particles and emphatic clitics.
11. On V as the head of S, see Jackendoff (1977). On the distinction of syntactic and semantic heads, see Abney (1986). Notice that in Bengali, I cannot take VP as its complement because this would predict that in *babulke mere-o-chilam* (Babul[obj] beat-too-have [I]) 'I/we have also beaten Babul' the clitic can focus on *babul*. This, however, is never possible. The clitic can only focus on the verb in such cases.

### References

- Abney, Steven  
1986 "Functional elements and licensing" [Paper presented at the GLOW conference Girona, Spain].  
1987 *The English noun phrase in its sentential aspect* [Dissertation, MIT, Cambridge, MA].
- Bashir, E.—M. Deshpande—P. Hook (eds.)  
1987 *Selected papers from SALA 7*. (Bloomington: Indiana University Linguistics Club).
- Bayer, Josef  
1988 "Directionality of government as a locality constraint for scope assignment" [Paper presented at the GLOW conference Budapest, Hungary].  
1990 "Interpretive islands: evidence for connectedness and global harmony in logical form", in: G. Grewendorf—W. Sternefeld (eds.).
- Chomsky, Noam  
1986 *Barriers*. (Cambridge, MA: MIT Press).
- Dasgupta, Probal  
1984 "Bangla emphaziers and anchors", *Indian Linguistics* 45: 102—117.  
1987 "Sentence particles in Bangla", in: E. Bashir—M. Deshpande—P. Hook (eds.), 49—75.  
[ms.]. "The word in Bangla", unpublished paper, Pune, Deccan College.
- Grewendorf, G.—W. Sternefeld (eds.)  
1990 *Scrambling and barriers*. (Amsterdam: Benjamins).
- Huang, J.—R. May (eds.)  
in press *Logical structure and linguistic structure*. (Dordrecht: Reidel).
- Jackendoff, Ray  
1977  *$\bar{X}$ -Syntax*. (Cambridge, MA: MIT Press).
- Kayne, Richard S.  
1981 "On certain differences between French and English", *Linguistic Inquiry* 12: 349—371.
- Koster, Jan  
1986 *Domains and dynasties*. (Dordrecht: Foris).
- Longobardi, Giuseppe  
in press "In defense of the 'correspondence hypothesis': island effects and parasitic constructions in logical form", in: J. Huang—R. May (eds.).

- May, Robert  
1977 *The grammar of quantification* [Dissertation, MIT, Cambridge, MA].
- Rooth, Mats  
1985 *Association with focus* [Dissertation, University of Massachusetts, Amherst].
- Ross, John R.  
1967 *Constraints on variables in syntax* [Dissertation, MIT, Cambridge, MA].
- Szabolcsi, Anna  
1984 "The possessor that ran away from home", *The Linguistic Review* 3:  
89–102.
- Zwicky, Arnold—Geoffrey Pullum  
1983 "Cliticization vs. inflection: English *n't*", *Language* 59: 502–513.