SPLITTING FRIENDS, WIVES, AND BOXES OF BOOKS*

TUE TRINH
University of Wisconsin - Milwaukee

1 Introduction

The copy theory of movement implicates a rule, Copy Deletion, which applies to delete the phonological content of the lower copy of the resulting chain. It remains to characterize the structural description of this rule. In Trinh (2009, 2010), I propose the following.

(1) The Edge Condition (EC)
For any chain (α, β) where α is the higher and β is the lower copy, phonological deletion of β requires that β end an XP

The phrase "end an XP" is to be understood as 'occupy the right edge of an XP,' or more precisely, 'have the rightmost morpheme co-incide with the rightmost morpheme of a non-projecting category.' As an example, consider the contrast between Vietnamese and German with respect to how topicalized verbs are pronounced (TOP = topic marker).[1]

(2) a. Đọc thì nó nên [$_{VP}$ *(đọc) sách]
read TOP he should read books
'He should read books'
b. Lesen sollte er [$_{VP}$ Bücher (*lesen)]
read should he books read
'He should read books'

*I thank Noam Chomsky, Danny Fox, Gennaro Chierchia, Irene Heim, Sabine Iatridou, David Pesetsky and the audiences in Berlin and Potsdam for valuable input.

[1]The translations of non-English sentences are intended to convey only truth-conditional content, abstracting from the interpretive effects of the information structure. I believe this is better than trying to capture these effects in the translations, resulting in sentences that may sound much less natural than the original sentences.

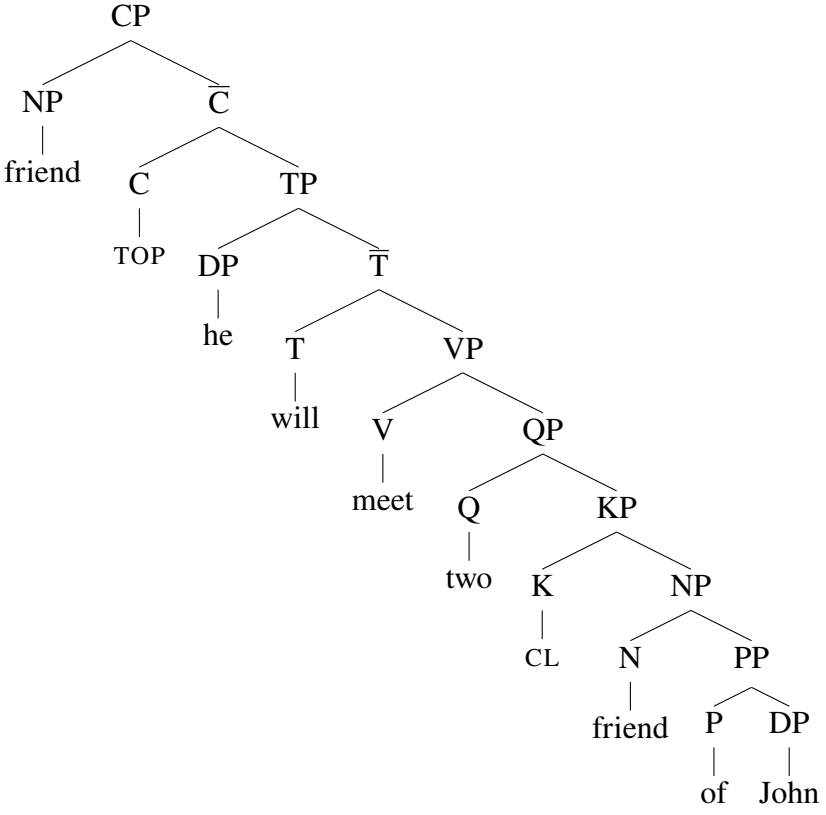
Suppose that both sentences in (2) involve movement of a bare V to [Spec,C] and that VP is head-initial in Vietnamese but head-final in German, EC makes the correct prediction that the moved V is pronounced twice, i.e. "doubled," in Vietnamese but not in German, assuming that Copy Deletion must apply when it can. Cross-linguistic variation with respect to the pronunciation of topicalized verbs make up most of my argument for EC in Trinh (2009, 2010). In section 5 of Trinh (2009), I argue that variation within Vietnamese with respect to the pronunciation of split NPs also supports EC. After the publication of that paper, I became aware of some additional facts about NP-Split in Vietnamese, which I then discussed in chapter 3 of my dissertation (Trinh, 2011). I believe these facts and my analysis of them suffice to warrant presentation in a small but self-contained contribution. The present squib is a long overdue result of that belief, and I am particularly happy to see it appear in this volume, as David Pesetsky's guidance and support were essential in all of my works on this topic.

2 Relational Nouns

I use the term "NP-split" to refer to the extraction, specifically topicalization, of an N(P) from a nominal phrase which includes a numeral, a classifier, and possible other elements. I will begin with the discussion of topicalized relational nouns, for example *vợ* 'wife' or *bạn* 'friend,' as illustrated in (3).

(3) a. bạn thì nó sẽ gặp hai người *(bạn) của John
friend TOP he will meet two CL *(friend) of John
'He will meet two friends of John'

b.

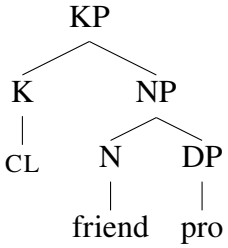


Parsing (3a) as (3b) accounts for the obligatory doubling of *bạn* 'friend' in the sentence, as its lower occurence does not end an XP in the sense specified above. Now it turns out that doubling of a fronted relational noun is not obligatory, but optional, if the noun has no complement.

(4) bạn thì nó sẽ gặp hai người (bạn)
friend TOP he will meet two CL (friend)
'He will meet two friends'

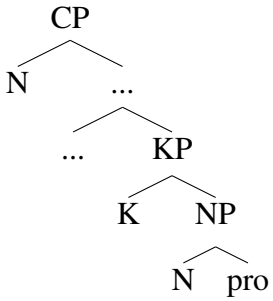
Let us derive this fact in the following way. Suppose that *bạn* 'friend' actually has a complement in (4), a silent pronoun, and consequently, that KP has the structure in (5).

(5)

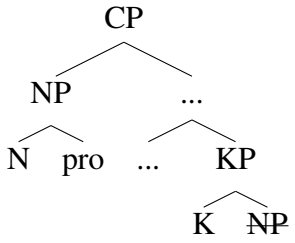


As **pro** is silent, topicalization of N will fill [Spec,C] with the same phonological material as topicalization of NP. Given EC, however, we expect that the phonological material at the base position will differ between these two cases: fronting N bleeds, while fronting NP feeds, Copy Deletion. Under the plausible assumption that fronting of N and fronting of NP are both possible, doubling of the topic in (4) follows.

(6) a.



b.



Are there independent reasons to say that (5) is the correct analysis of KP? Yes. Standard semantics of nouns and classifiers actually requires that *pro* be there.

(7) a. $[\![\text{bạn}]\!] = \lambda x \in D_e \,.\, \lambda y \in D_e \,.\, y$ is friends with x
b. $[\![\text{người}]\!] = \lambda P \in D_{<e,t>} \,.\, P \cap \{x \mid x \text{ is an atomic individual}\}$

The classifier maps a predicate P to a set of atomic individuals falling under P, and is therefore of type <<*e*,*t*>,<*e*,*t*>> (Chierchia, 1998). The relational noun denotes a relation and is therefore of type <*e*,<*e*,*t*>>. This means merging the noun directly with the classifier will result in a type mismatch. The mismatch can be circumvented by first merging the noun with *pro*, which is of type *e*, and then merging the complex NP, which is of type <*e*,*t*>, with the classifier, as shown in (5).

Note that although the topic constituents in (6a) and (6b) sound the same, they do not mean the same: the bare noun denotes a relation, the NP a predicate. Can an experiment be designed to show this difference? Since topicalization alters the information structure and not the propositional

content of sentences, judgements will have to pertain to discourse coherence, not truth condition. Consider the paradigm in (8), where capitalization indicates focal stress.

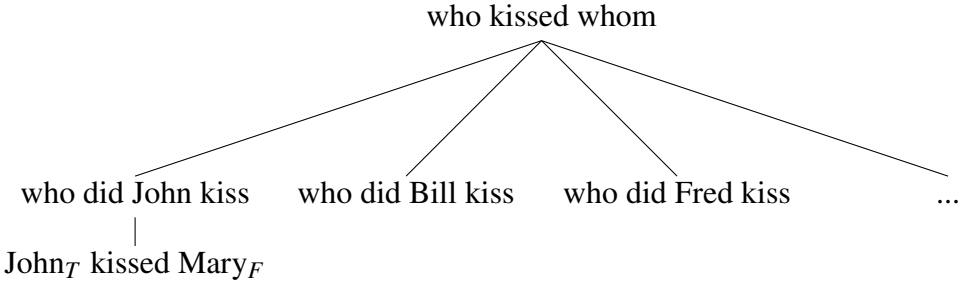
(8) a. Vợ thì nó gặp HAI người. Bạn thì nó gặp BA người.
wife TOP he met two CL friend TOP he met three CL
'He met two wives and three friends'
b. Vợ thì nó gặp HAI người vợ. Bạn thì nó gặp BA người bạn.
wife TOP he met two CL wife friend TOP he met three CL friend
'He met two wives and three friends'
c. #Vợ thì nó gặp HAI người. Bạn thì nó gặp BA người bạn.
wife TOP he met two CL friend TOP he met three CL friend
('He met two wives and three friends')

The base position of the topicalized constituent is empty in both of the sentences of (8a), in none of the sentences of (8b), and in the first but not the second sentence of (8c). It is (8c) which stands out as being odd. My account of this oddness will rely on the notion of "topic value," as defined in Büring (1999, 2003). In these works, Büring proposes that sentences containing a focus and a contrastive topic, such as those in (8), have a third semantic value, the topic value, in addition to the focus and the ordinary value. By definition, the topic value of S, $[\![S]\!]^t$, is the set of focus values $[\![S']\!]^f$ where S′ is derived from S by replacing the topic constituent in S with an expression of the same type. The focus value of S is computed as defined in Rooth (1985, 1992, 1996): it is a set of ordinary values $[\![S']\!]^o$ where S′ is derived from S by replacing the focused constituent in S with an expression of the same type. An illustration is given in (9), with the subscripts T and F marking the topic and the focus of the sentence, respectively.[2]

(9) a. $[\![\text{John}_T \text{ kissed Mary}_F]\!]^o = \text{John kissed Mary}$
b. $[\![\text{John}_T \text{ kissed Mary}_F]\!]^f = \{\text{John kissed } y \mid y \in D_e\}$
c. $[\![\text{John}_T \text{ kissed Mary}_F]\!]^t = \{\{x \text{ kissed } y \mid y \in D_e\} \mid x \in D_e\}$

Given the proposition-set theory of questions (cf. Hamblin, 1973), (9a) is the question 'who did John kiss' and (9b) the set containing such questions as 'who did John kiss', 'who did Bill kiss', 'who did Fred kiss', etc. These questions, in turn, can be seen as subquestions of a "superquestion," namely 'who kissed whom.' This is represented in the following "discourse tree," or "d-tree."

(10)

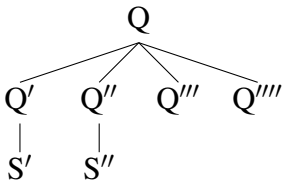


I assume, following Büring (1999, 2003), that a sentence is felicitous only if it can be a node in an available d-tree, and that the following condition holds.

[2]Topic is usually marked in English by the L-H* pitch contour (Pierrehumbert, 1980), also called the "B-accent" (Jackendoff, 1972).

(11) CT-Congruence
A sentence S containing a topic and a focus can be a node in a d-tree D only if the question Q dominating S in D, and all of Q's sisters, are elements of $[\![S]\!]^t$.

This means, for (12), that it has to be the case that $[\![S']\!]^t = [\![S'']\!]^t = \{Q', Q'', Q''', Q''''\}$.

(12)


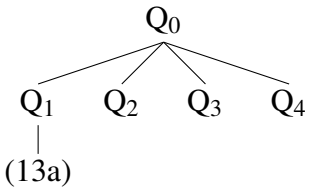
Now let us come back to the oddness of (8c). The first sentence in this sequence does not show doubling of the fronted relational noun *vợ* 'wife'. It thus instantiates (6b) and has the focus and topic markings shown in (13a). The topic value of (13a) is the question set in (13b).

(13) a. he met two$_F$ [wife of pro]$_T$
b. $\{\{$he met *n P* $\mid n \in \mathbb{N}\} \mid P \in D_{<e,t>}\}$ = {how many wives of pro did he meet, how many friends of pro did he meet, how many linguistics students did he meet, how many female democrats did he meet}

The second sentence in (8c) does show doubling of the fronted relational noun *bạn* 'friend'. This means it instantiates (6a). Its information structure and topic value are given in (14a) and (14b).

(14) a. he met three$_F$ [friend]$_T$ of pro
b. $\{\{$he met *n R* of pro $\mid n \in \mathbb{N}\} \mid R \in D_{<e,et>}\}$ = {how many friends of pro did he meet, how many wives of pro did he meet, how many children of pro did he meet, how many siblings of pro did he meet}

Now suppose that the utterance of (13a) does two things: (i) it establishes the d-tree in (15) with Q_1, ..., Q_4 representing the questions in the topic value of this sentence, i.e. the elements of (13b); and (ii) it reduces the set of "available d-trees" to those which are extensions of (15), i.e. which are derivable from (15) by plugging in the answers to Q_2, Q_3 or Q_4.

(15)


It then follows that the second sentence of (8c), (14a), is infelicitous, as it cannot be a node in any of the available d-trees. Specifically, its being daughter of any of the subquestions of Q_0, including the question 'how many friends of pro did he meet', will incur violation of CT-Congruence, because $[\![(14a)]\!]^t \neq \{Q_1, Q_2, Q_3, Q_4\}$.

We have thus explained the oddness of (8c). What about the acceptability of (8a) and (8b)? The reader is invited to verify for herself that in these sequences, the first and the second sentence have identical topic value and therefore every sentence can be a node in an available d-tree.

Here is a potential counterexample to what I have said so far.

(16) Vợ thì nó gặp HAI người vợ. Bạn thì nó gặp BA người.
wife TOP he met two CL wife friend TOP he met three CL
('He met two wives and three friends')

I predict the sequence in (16) to be infelicitous: the first sentence shows doubling, which indicates the topic is N; the second does not show doubling, which indicates the topic is [N pro]. It follows that these sentences do not have the same topic value, and should be infelicitous. The problem is that they are perfectly acceptable. My solution is to say that the lower copy of the fronted noun **bạn** 'friend' is eliminated from the phonetic representation of the sentence not by Copy Deletion, but by ellipsis. I propose that it is also ellipsis which takes place in the second sentence of (17), where not only the lower copy of the fronted noun but also the classifier preceding it is elided.

(17) Vợ thì nó gặp HAI người vợ. Bạn thì nó gặp BA.
wife TOP he met two CL wife friend TOP he met three
'He met two wives and three friends'

Evidence that (16) and (17) involve ellipsis is the fact that switching the order of the sentences in (16) and (17) leads to deviance in both cases.

(18) a. *Bạn thì nó gặp BA người. Vợ thì nó gặp HAI người vợ.
friend TOP he met three CL wife TOP he met two CL wife
('He met three friends and two wives')
b. *Bạn thì nó gặp BA. Vợ thì nó gặp HAI người vợ.
friend TOP he met three wife TOP he met two CL wife
('He met three friends and two wives')

Also, it should be noted that when we conjoin the two clauses in (18a) and (18b), making the sequence one big sentence, acceptability increases greatly.

(19) a. Bạn thì nó gặp BA người nhưng vợ thì nó chỉ gặp HAI người vợ.
friend TOP he met three CL but wife TOP he only met two CL wife
'He met three friends but only two wives'
b. Bạn thì nó gặp BA nhưng vợ thì nó chỉ gặp HAI người vợ.
friend TOP he met three but wife TOP he only met two CL wife
'He met three friends but only two wives'

As cataphoric ellipsis is possible only internal to one sentence, the fact that both sentences in (19) are grammatical is additional evidence that ellipsis is involved in apparent counterexamples to our prediction.

3 Measure Words

In Vietnamese, container words such as **thùng** 'box' or **túi** 'bag' are systematically ambiguous between a "noun reading," exemplified in (20a), and a "measure word reading," exemplified in (20b). I will gloss the container noun **thùng** 'box' in its measure word reading as "MW_{box}."

(20) a. John mua hai cái thùng
John bought two CL box
'John bought two boxes'

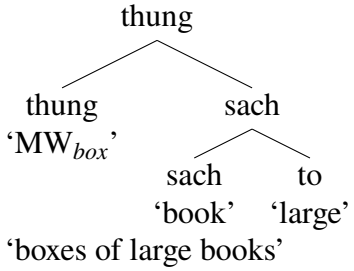
b. John mua hai thùng sách
John bought two MW_{box} book
'John bought two boxes of books'

Syntactically, a classifier intervenes between the container word and the numeral in (20a) but not in (20b). Semantically, (20a) is true only if John ended up in possession of two physical boxes, while (20b) can be true even if John paid money only for the books that fit, or would fit, into two boxes but not for any box. I propose (21) as the meaning of MW_{box}.

(21) $[\![\text{MW}_{box}]\!] = \lambda P \in D_{<e,t>} . \lambda x \in D_e . x$ is a box-load of things that are P

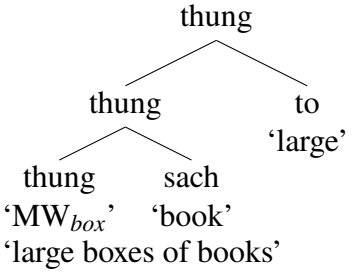
Thus, measure words resemble classifiers in being functions of type <<*e*,*t*>,<*e*,*t*>>. However, they are, in a sense, more substantive than classifiers. For example, the classifier **quyen** in effect maps books to books, while the measure word **thung** would map books to boxes of books. This difference has interesting repercussions for modification. Let us say that an adjective such as **to** 'large' denote the set of large entities, i.e. a predicate of type <*e*,*t*>. Consider (22a) and (22b).[3]

(22) a.



'boxes of large books'

b.



'large boxes of books'

In (22a), the adjective **to** 'large' modifies **sach**, which denotes the set of books, and the whole phrase denotes boxes of large books. In (22b), the same adjective modifies **thung sach**, which denotes not the set of books, but the set of box-sized quantities of books. Assuming that boxes come in different sizes, the set of box-sized quantities of books will include those the size of a small box and those the size of a large box. In other word, $[\![\textit{thung sach}]\!]$ contains small quantities and large quantities of books. Intersecting $[\![\textit{thung sach}]\!]$ with $[\![\textit{to}]\!]$, which is what the interpretation of (22b) comes down to, will yield a subset of $[\![\textit{thung sach}]\!]$ which contains only the large quantities in $[\![\textit{thung sach}]\!]$. This is the set of large boxes of books.

We predict, then, that (23) is ambiguous between 'John will buy two boxes of large books' and 'John will buy two large boxes of books.' This prediction is correct.

(23) John mua hai thùng sách to
John bought two MW_{box} book large
'John bought two large boxes of books / John bought two boxes of large books'

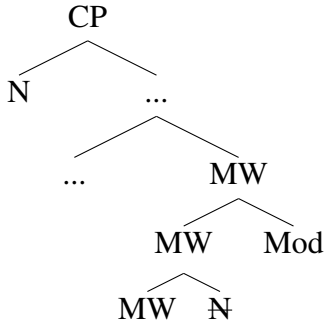
Now let us ask what prediction can be made about NP-split constructions in which the split nominal complex contains a measure word and a modifier. Consider the following sentence.

(24) Sách thì John mua hai thùng to
book TOP John bought two MW_{box} large
'John bought two large boxes of books / *John bought two boxes of large books'

[3]Note that we assume, for now, that modifiers do not project, and that heads which denote functions that take their sisters as arguments do project.

The observation is that (24) is unambiguous: it has the reading where the adjective modifies 'boxes of books,' but not the one where it modifies 'books.' It turns out that this is exactly what we predict. Here is how. The fact that there is no doubling of the fronted noun in (24) means that Copy Deletion must have applied, which means that the lower copy of **sach** 'book' must be at the right edge of an XP. which means that **sach** 'book' in (24) is fronted from (22b), not (22a). In other word, the analysis of (24) must be (25).

(25)

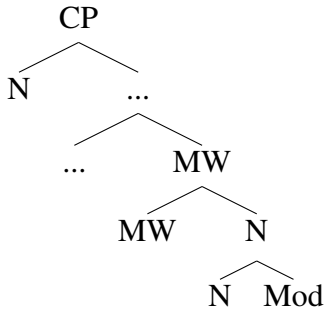


We make the correct prediction, then, that (24) must mean John will buy two large boxes of books and cannot mean John will buy two boxes of large books, since only the former meaning can be computed from (25). Now let us consider the doubling variant of (24), i.e. the sentence which differs minimally from (24) in that the topicalized noun is doubled.

(26) Sách thì John mua hai thùng sách to
book TOP John bought two MW_{box} book large
'John bought two large boxes of books / John bought two boxes of large books'

As the translation shows, (26) is ambiguous in exactly the same way as (23) is. This fact is not what we predict. As there is doubling of the topic noun **sach** 'book' in (26), Copy Deletion must have been blocked, which means that the base position of the chain created by topicalization must not be XP-final. The conclusion, then, is that **sach** 'book' must have been fronted from (22a), i.e. that (26) must have the structure in (27).

(27)



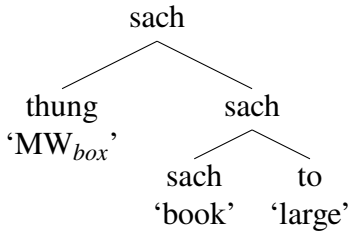
However, that means (26) can only be read as 'John will buy two boxes of large books.' In other word, we predict, incorrectly, that (26) is not ambiguous.

One solution to this problem is to stipulate that in case the lower copy is the complement of a measure word, Copy Deletion can but does not have to apply. This enables us to say that the semantic

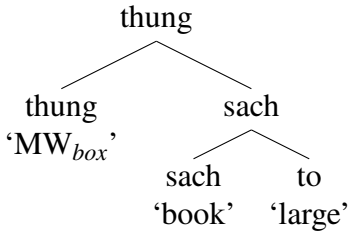
ambiguity of (26) is due to the fact that it is structurally ambiguous: it can be parsed as (25) with Copy Deletion not applying, or as (27) where Copy Deletion cannot apply. This solution, however, is obviously ad hoc and I will not adopt it.

Another solution, which was suggested to me by David Pesetsky and which I will adopt, is to assume that when a measure word and noun merge, projection is free choice: either the measure word or the noun can project. Thus, the string **thung sach to** 'MW$_{box}$ book large' would have four possible analyses.

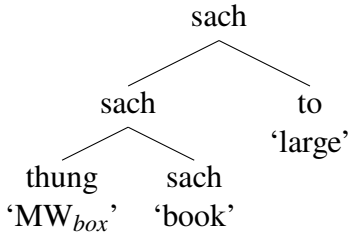
(28) a.



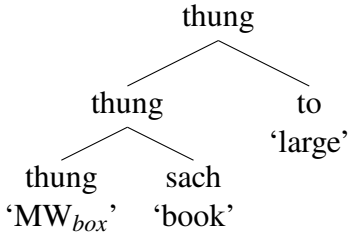
b.



(29) a.



b.



Among the four structures in (28) and (29), only one, namely (29b), is such that fronting **sach** 'book' from it will not result in doubling. As (29b) denotes 'large box of books,' we predict, correctly, that (24) is unambiguous, i.e. that it has to mean 'John will buy two large boxes of books': the noun **sach** 'book' in that sentence must have been fronted from (29b), since the lower copy of **sach** 'book' has been deleted and (29b) is the only structure in which this copy is XP-final. Fronting **sach** 'book' from any of the remaining structures will force doubling of the fronted constituent, since in none of these structures is **sach** 'book', in its base position, XP-final. We predict, again correctly, that (26) is ambiguous: the noun **sach** 'book' in it could be fronted from one of the structures in (28), in which case the sentence would mean 'John will buy two large boxes of books,' or it could be fronted from (29a), in which case the sentence would mean 'John will buy two boxes of large books.'

References

Büring, Daniel. 1999. Topic. In Focus — linguistic, cognitive, and computational perspectives, ed. Peter Bosch and Rob van der Sandt, 142–165. Cambridge University Press.

Büring, Daniel. 2003. On d-trees, beans, and b-accents. Linguistics and Philosophy 26:511–545.

Chierchia, Gennaro. 1998. Reference to kinds across languages. Natural Language Semantics 6:339–405.

Hamblin, Charles Leonard. 1973. Questions in Montague English. Foundations of Language 10:41–53.

Jackendoff, Ray. 1972. Semantic Interpretation in Generative Grammar. MIT Press.

Pierrehumbert, Janet. 1980. The phonetics and phonology of english intonation. Doctoral Dissertation, MIT.
Rooth, Mats. 1985. Association with focus. Doctoral Dissertation, UMASS-Amherst.
Rooth, Mats. 1992. A theory of focus interpretation. Natural Language Semantics 1:75–116.
Rooth, Mats. 1996. Focus. In The handbook of contemporary semantic theory, ed. Shalom Lappin, 271–297. Oxford: Blackwell Publishers.
Trinh, Tue. 2009. A constraint on copy deletion. Theoretical Linguistics 35:183–227.
Trinh, Tue. 2010. Edges and linearization: A reply. Theoretical Linguistics 36:93–110.
Trinh, Tue. 2011. Edges and linearization - an investigation into the pronunciation of chains. Doctoral Dissertation, Massachusetts Institute of Technology.