The Relationship between the Pragmatics of Preposed Objects and the Decline in Topicalization in the History of English[*]

Laura Whitton

1 Introduction

In this study, object topicalization is defined purely syntactically as an operation that preposes an object into clause-initial position. Note that in earlier stages of English, this preposing could trigger subject-verb inversion as in Example (1) or not as in Example (2).

(1) *Al this* **say I** not oonly for the but for other that in tyme comynge mowe atteyne to suche a forme of lyuynge. (CMAELR4,4.104)
(2) *Al this* **I saye** for that I wolde thou louedist silence and litel speche. (CMAELR4,5.130)

Topicalization is distinct from left dislocation in that it does not have a coreferential pronoun in canonical position and also has a different discourse function. (For the pragmatics of left dislocation see Prince, 1997.) Topicalization can also be differentiated from Focus Movement in that the tonic stress does not fall on the preposed constituent.

Topicalization remains possible in Modern English, though concomitant movement of the verb is ungrammatical, as demonstrated in (3) and (4):

(3) I love pit bulls, but poodles I hate.
(4) I love pit bulls, *but poodles hate I.

Speyer (to appear) presents a quantitative diachronic account of the decline in the frequency of object topicalization from early Middle English through

---

Early Modern English reproduced as Graph A. Since non-canonical word orders usually serve an information packaging function, the decline in usage of topicalization could be related to a change in what the construction conveys about information status of the entities in the sentence and their relationship to the surrounding context.
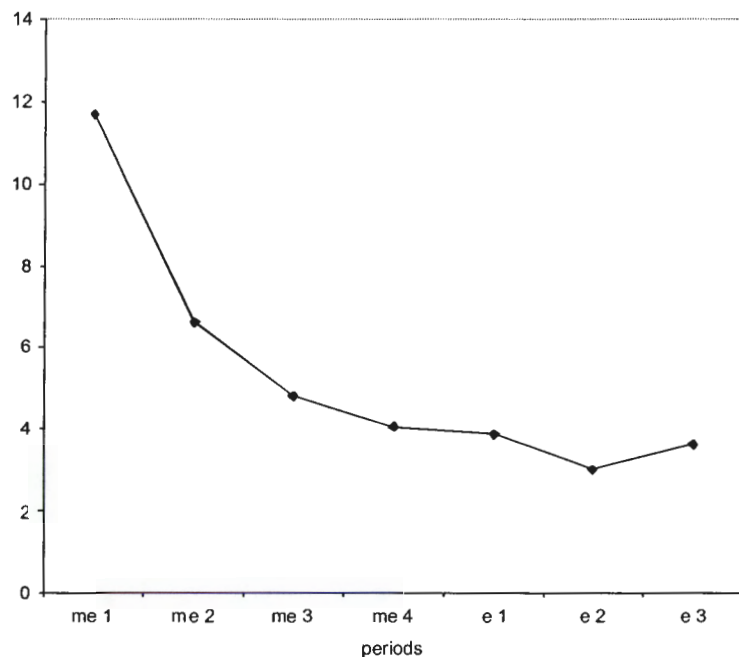
Graph A: The Decline in Percentage of Direct Objects Topicalized in Middle (me1–me4) and Early Modern (e1-e3) English (Speyer, to appear)

The pragmatic conditions under which topicalization occurs in English will be examined below. A diachronic quantitative analysis of these conditions follows, which demonstrates that no change in the discourse constraints on object preposing did in fact take place in the evolution from Middle to Modern English which could explain the pattern in Graph A.

2 The Discourse Linking Function of Topicalization

Early accounts of topicalization relied on the general observation that new information was precluded from the topic position and the intuitive notion that the preposed entities were what the sentence was "about." Prince, however, provides counterexamples to such theories and states that for an NP to be felicitously preposed it has to represent an entity that was either already evoked in the discourse or else was inferentially related, via a salient-set relation, to an entity already evoked.

A formal account whose explanatory power covers both the evoked and salient-set types and which also further specified the nature of the set relationship was presented in Ward (1985) and Birner and Ward (1998). Ward (1985) claims that a partially ordered set relation analysis, as applied to scalar implicature in Hirschberg (1981), supports all of the topicalizations in a large corpus of data. Under Ward's model, "preposing marks the referent or denotation of the preposed constituent as a Backward Looking Center (Joshi and Weinstein, 1981; Grosz, Joshi and Weinstein, 1983)" and "marks an Open Proposition as salient in the discourse." Furthermore, the preposed constituent must be related via a salient partially ordered set relation to one or more entities already evoked in the discourse model.

Formally, partially ordered sets are defined by a partial ordering R on some set of referents, b, such that for all b-1, b-2, and b-3 that are elements of b, R is either reflexive, transitive and antisymmetric or irreflexive, transitive and asymmetric.

REFLEXIVE:	b-1 R b-1
TRANSITIVE:	(b-1 R b-2 and b-2 R b-3) → (b-1 R b-3)
ANTISYMMETRIC:	(b-1 R b-2 and b-2 R b-1) → (b-1 = b-2)
IRREFLEXIVE:	b-1 does not R b-1
TRANSITIVE:	(b-1 R b-2 and b-2 R b-3) → (b-1 R b-3)
ASYMMETRIC:	(b-1 R b-2) → (b-2 does not R b-1))

An example of the second definition is *is-greater-than* and an example of the first is *is-greater-than-or-equal-to*. Other poset relations include *is-a-part-of*, *is-a-subtype-of*, and *is-a-member-of* (Ward, 1985:65). Under this theory, relations that do not fit the poset definition, e.g. relations that are not transitive or that are symmetric, are disallowed in felicitous topicalization. For example, the relation of functional dependence, which is sufficient to

support the use of a definite article, as shown in (5), is insufficient to support topicalization, as shown in (6):

(5)   John went into a restaurant and he asked for the menu. (Ward and Prince's 17a)

(6)   #John went into a restaurant and the menu he asked for.   (Ward and Prince's 17b)

Ward and Prince claim to demonstrate the lack of transitivity of functional dependence with the examples in (7) and (8).

(7)   We ate in a terrible French restaurant last night.
      #The cork was green. (Ward and Prince's 18a)

(8)   We ate in a terrible French restaurant last night. The wine was awful. The cork was green. (Ward and Prince's 18b)

Thus Ward and Prince conclude that those relations that must obtain between an entity and its context for felicitous topicalization are just those relations that form a natural class on independent grounds, namely poset relations.

The second type of discourse linking relation analyzed in this study is identity. Identity is technically a poset relation. However, as it is only thematic and not rhematic in the sense of Vallduví and Vilkuna (1998), it is analyzed separately here. As it is not quantificational in nature, it is possible that this difference would be a relevant factor in a syntactic/pragmatic change.

Identity relations can be that of pure identity (9) or synonym/restatement (10). Bare demonstratives, as in (11), are the most common case of pure identity. Object pronouns were excluded because of their clitic status in earlier stages of English. The synonym/restatement cases often reformulate an evoked entity or proposition, marking the fact that their referent has already been evoked with a demonstrative, such as the topicalized *this same rule* in (10). Cases of discourse segment reference in the sense of Webber (1998) were included in identity cases.

(9)   'Well,' seyde the damesell, 'ye ar nat wyse to kepe the swerde fro me, for ye shall sle with that swerde the beste frende that ye have and the man that ye moste love in the worlde, and that swerde shall be youre destruccion.'
      'I shall take the aventure' seyde Balyn, 'that God woll ordayne for me.

> But *the swerde* ye shall nat have at thys tyme, by the feythe of my body!' (CMMALORY,47.1562)

(10)   for <u>thou shalt slepe no more of al day</u>. *This same rule* shalt thou kepe from Estern vnto the kalendes of Nouember, saaf that thou shalt slepe after mete afore the houre of noon. (CMAELR4,6.165)

(11)   Thereupon he devised <u>a little wheel</u>, with all the Capital Romane letters made upon a paper to wrap round about it, and fitted it to turn in little a round box, which had a hole so made in the side of it, that only one letter might be seen to peep out at once; *This* he brought to the childe, & showed him onely the letter O, and told him what it was; (HOOLE,9.55)

Since most accounts of preposing in English rely on some notion of a "discourse-linking function" with the previous text (Birner and Ward, 1998) or marking a Backward Looking Center (Ward 1985), the relationship between the preposing and the *subsequent* discourse is often ignored. However, the corpus contained several examples of identity relations that were cataphoric, such as in (12). These were counted as identity cases:

(12)   But *this effect* it wrought, all the Lords were of opinion that his Highness wordes and Articles must bee made good; that the oath by the Councell must bee taken; (CONWAY1,2.3,156.16)

A thorough description of the environments in which topicalizations occur should account for such tokens, since they are quite common in written and spoken Modern English.

Similarly, cases were coded as posets when there was an explicit entity in the subsequent discourse that was in a poset relation with the preposed entity. As an illustration, the discourse initial example from Modern English in (13) would be coded as a poset:

(13) **Brains you're born with**. A great body you have to work for. (Brooke Shields, in a health club commercial.) (= Ward and Prince's 7a)

There are certain cases that can be viewed as both identity and poset. These are cases such as (14), where *that* is clearly co-referential with *a texte in englysshe* and in that respect could be coded as a bare demonstrative identity case. However, it is clear that the sentence with the preposed *that* is meant to render salient its alternative, a text in Latin, in the set of {Texts in Different Languages}.

(14) but bycause this texte of sayncte Paule is in latyn, and husbandes commonely can but lyttell laten, I fere leaste they can-not vnderstande it. And thoughe it were declared ones or twyse to theym, that they wolde forgette it: Wherfore I shall shewe to theym a texte in englysshe, and *that* **they maye well vnderstande**, and that is this, Eate within thy tedure. (FITZH,99.306)

Because the alternatives in the set are clearly activated at the time of the preposing and the alternatives explicitly appear in the passage, this (and other similar examples) were coded as posets.

Tokens like those in (15) and (16) support the idea that the discourse constraints on topicalization have narrowed over time, since their Modern English equivalents given in (15') and (16') are not felicitous today:

(15) vpon bed he lay neuyr;
      ***hosen and schon* wered he nowt**; (CMCAPCHR,56.721)
(15') upon a bed he never laid
      ***socks and shoes* he never wore**
(16) Here deieth Dauid, +te son of Jesse.
      He was born of +te tribe of Juda in +te cite' of Bethlem, fayre in nature, wise in prophecye, both kyng and prophete.
      ***Kyngis* he ouyrcam with uictorye;**
      ***psalmes* he sang with melodie;**
      ***bestes* he killid** , and Goly the grete geaunt.
      Euyr he dred God.
      Cristis natiuite', his baptem, his passion, resurreccion, ascencion, his comyng to +te dome, ful openly in his psalmes he teld.
      ***His fadir scheep* kepte he ful mekly**; (CMCAPCHR,32.13)
(16') Here died David, the son of Jesse...
      ***Kings* he overcame with victory**
      ***Psalmes* he sang with melody**
      ***Beasts* he killed...** etc.

Since the types of sentences like those in (15) and (16) do not conform to current theories of the pragmatics of English, it seemed likely that their disappearance was involved in the decline in topicalization. Of course, if the discourse situations that allowed topicalization became more restricted, speakers/writers would not topicalize as frequently.

3 Results

The data consist of approximately 475 sentences with preposed objects from the Penn Helsinki Parsed Corpus of Middle English (Kroch and Taylor, 2000), the Penn Helsinki Parsed Corpus of Early Modern English (Kroch and Santorini, in preparation), the Brown Corpus (Francis and Kucera, 1982) and the British National Corpus. Matrix clauses with topicalized objects were extracted[1] using CorpusSearch (Randall, 2000) and coded for the relationship of the preposed entity, either "poset," "identity," or, if neither, "residual." The results are presented in Graph B.
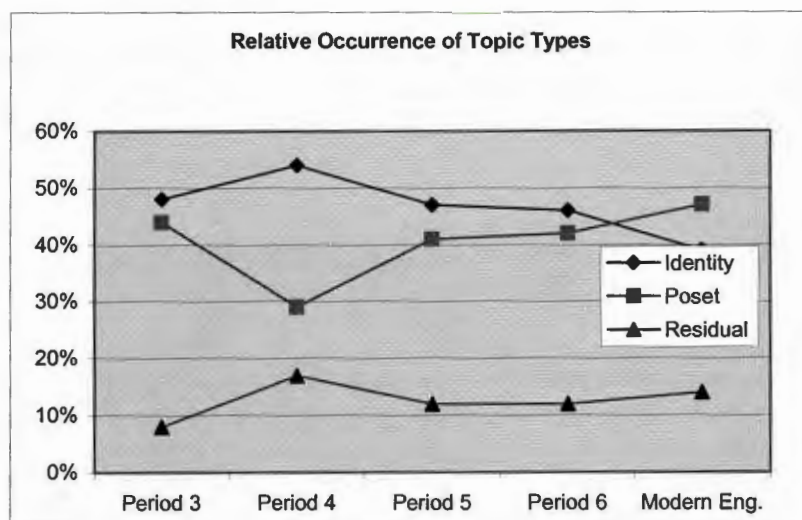
Graph B: Relative Occurrence of Topic Types

---

[1]The data for this study were also used for Whitton (2003), which analyzed both the loss of topicalization and V2. As such, sentences with main verb *be* were excluded because previous work (Johnson and Whitton 2002) demonstrated that inversion behaves differently in copular sentences. Fronted pronominal objects were also excluded due to their clitic status in earlier stages of English. Quotative sentences were removed from the data due to the existence of quotative inversion in Modern English. Sentences with preposed negative phrases were also excluded since

Relative Occurrence of Topic Types

	Period 3 1300-1399		Period 4 1400-1499		Period 5 1500-1611		Period 6 1612-1712		Modern English	
	N	%	N	%	N	%	N	%	N	%
I	45/93	48	50/93	54	43/92	47	44/95	46	27/72	38
P	41/93	44	27/93	29	38/92	41	40/95	42	35/72	49
R	7/93	8	16/93	17	11/92	12	11/95	12	10/72	14

Although Period 4 behaves somewhat differently, the relative stability of the topic types during these periods suggested that there had been no change in the function of topicalization during Middle and Early Modern English. But the existence of the residual cases, which presumably do not exist in today's language, suggested that a narrowing in topicalization did in fact occur. However, when data from Modern English were examined, the topic types were distributed almost exactly as in the other periods, with the residual cases continuing to account for slightly more than 10% of the data. Furthermore, a chi square test shows that the differences between the distribution of identity, poset and "other" cases from period to period are not significant ($p \leq .1$)

The data from Modern English show that 38% of the topicalized object sentences examined had a preposed entity in an identity relation with its surrounding discourse. We must note, however, that in Ward's corpus of 409 tokens of Modern English NP topicalization, only 17 (4%) are coded as involving the identity relation. But this difference can be attributed to the fact that Ward's coding, as mentioned above, treats only a certain case of preposing as an identity relation: those in which NPs perform a bridging relation in the discourse by providing an "additional description or summary of the entity to which it refers" (Ward, 1985:209) as in (17):

(17) Cohen presents a number of arguments against assuming innate universals. He points out analogies between language acquisition and scientific discovery, concluding that by parity of reasoning, if the assumption of innate linguistic universals is required for the first, then

---

they also still invert. Biblical texts, tokens from legal documents of the form "This witnesseth:_____" and the first lines of letters were also excluded.

some analogous assumption is required for the second. **This conclusion he takes to be more or less a reductio ad absurdum.** To account for scientific discovery, he argues, it suffices to postulate the "general capacity for eliminative induction." (Chomsky, N. 1975. *Reflections on Language*. 1975:205) (=Ward's 352).

Based on such examples and the infelicity of constructed examples as in (18), Ward concludes that the only "identity" topicalizations that are possible include reformulation and/or additional modifiers. Ward goes on to note that in addition to being rare, these type of summary/description identity cases were found primarily in formal, written registers (15 of the 17 cases) and that they are less felicitous in (constructed) less formal and less abstract scenarios:

(18) Wherever I go in Philadelphia, it seems I see something terrible happen. Just yesterday, I saw a guy get mugged and beaten up. #**This gross crime I saw while I was walking down Spruce Street.** (= Ward's 360)

Ward hypothesizes that topicalization is becoming more and more limited to non-identity scalar relations, and that this development may be related to the increasingly rigid word order which has characterized the development of English. My data, however, do not support Ward's hypothesis. Although posets do overtake identity cases in frequency by Modern English in my corpus, the difference between Early Modern English and Modern English in the relative occurrence of poset and identity cases is not significant at (p>.05).

4 Other Pragmatic Factors

4.1 Functional Dependence

The possibility remains that there has been some change in the pragmatics of topicalization that has contributed to its decreasing frequency, but that does not correspond to the categories examined. Examples like (19) from Early Modern English raise the question of whether functional dependence, coded herein as "residual" used to be possible as a linking relation for topics and has been lost.

(19) The Christmas following was an ill Feast to Edric, of whose Treason, the King having now made use as much as serv'd his turn, and fearing himself to be the next betray'd, caus'd him to be slain at London in the Palace, thrown over the City Wall, and there to lie unburied; **the head of Edric fixt on a pole, he commanded to be set on the highest Tower of London** (MILTON,X,276.148)

However, in (20), taken from Modern English, *the titles* are functionally dependent on "books" (not explicitly mentioned in the discourse until the subsequent synonym *works)* which are in turn functionally dependent on *bookcases*.

(20) He tried the doors of the bookcase. Locked. **The titles he could read easily through the criss-crossed wires**: works on theology, astral physics, history, biology, political science. No poetry. No novels. (Brown Corpus, ck13.mrg)

Tokens like (20) and other residual cases pose problems for theories that rely on the notion that only poset relations license topicalization. If further investigation reveals a significant number of such cases in corpora of Modern English, the category of functional dependence, which intuitively overlaps with part/whole in some instances, will need to be carefully defined before the poset theory can be reliably tested.

4.2 The Role of the Salient Open Proposition

Other historical examples suggest that no change in the pragmatics of preposing has occurred that involves the linking relation of the preposed NP, but there may have been one involving a change in the other function of topicalization: marking an open proposition as salient in the discourse. After introducing syntactic marking of OPs in English generally, I will discuss the role of presupposed open propositions in topicalization.

As described in Prince (1986), two types of propositions may be marked as presupposed: whole propositions and propositions containing a variable, so-called "open propositions". The latter can be marked by stress or by syntactic form, as demonstrated by Prince's examples given in (21):

(21)   a.   She gave the SHIRT to Harry
         b.   It was the SHIRT that she gave to Harry.
         c.   She gave X to Harry.

The stressed constituent in both (a) and (b) represents the variable in the OP given in (c). In (b), the it-cleft construction obligatorily marks syntactically (in addition to prosodically) that the OP in (c) is salient. Other syntactic marking of salient OPs can be found in questions, wh-clefts, and focus movement.

In her comparison of left-dislocation and topicalization, Prince (1997) states that the poset inference that the fronted element in triggers is common to both constructions, while the second element in the "Double Discourse Function of Topicalization" given in (22) is unique to topicalization:

(22) First, if the entity evoked by the leftmost NP represents an element of some salient set, make the set-membership explicit. Then, in all cases, the open proposition resulting from the replacement of the tonically stressed constituent (in the clause) with a variable is taken to represent information saliently and appropriately on the hearer's mind at that point in the discourse, the tonically stressed constituent representing the instantiation of the variable and the new information. (Prince, 1997:10)

An example of the application of this process follows (Prince's 12):

(23) She had an idea for a project. She's going to use three groups of mice. One, she'll feed them mouse chow, just the regular stuff they make for mice. Another, she'll feed them veggies. And **the third, she'll feed junk food.**

Open Proposition: She'll feed the third ($\in$ {the three groups of mice}) X.
Instantiation: X = junk food

To demonstrate that the set relation that must obtain in a topicalization is in fact independent of the OP requirement, Ward gives the following example:

(24)   A.   My new boyfriend really likes sports.
        B.   **#Well, badminton I played in high school**
               (Cf. Well, I played badminton in high school.)

Even though badminton is an element of the set "sports," the OP necessary to make B's response coherent is not salient in the discourse (Ward, 1985:98), as it would be if A had uttered something like "Did you play any sports growing up?" Hence, the topicalization in (24) is not felicitous.

Although a quantitative analysis of OPs was not conducted in this study, the requirement that topicalizations mark a salient OP is not met by several historical examples, such as the following example from Early Modern English (25)[2], suggesting that this constraint may have developed recently.

(25) ffor this is of all certentye, ether it is for Irelande or Englande, if for Englande, then are wee to doubt the taking in of theArmy of the Lowe Contreys to be transported, which (as it falleth out) need not feare ffraunce if the Truce be made as all the world sayes it is, and as shrewd circumstances discover: ffor wee see the Car=all= hath lost Amyens, followes not Count Mawryce, but keepes his fforces together and now soddenly drawes them downe to the sea syde. ***This* her majesty Requires you to lay feelinglie before the king,** as Reasones sufficient to dispence with her Revocacon. (RCECIL, 304.19)

However, tokens like (26) from Modern English demonstrate that topicalizations continue to exist in which no salient open proposition is present, again suggesting, albeit inconclusively, that there has been no change in the requirement that topicalization mark a salient OP.

(26) At five o'clock that night it was already dark, and behind my closed door I was dressing as carefully as a groom. I wore a new double-breasted brown worsted suit with a faint herringbone design and wide lapels like a devil's ears. My camp-made leather wallet, bulky with twisted, raised stitches around the edges, I stuffed with money **I had been saving.** Hatless, in an overcoat of rough blue wool, I was given a proud farewell by my mother and father, and I set out into the strangely still streets of Brooklyn. (Brown Corpus ck 29)

5 Summary

This study investigated the possibility that the decline in topicalization between the Middle English and Modern English periods could be explained by the fact that topicalization was possible at earlier stages of English in contexts which no longer enable felicitous topicalizations. This investigation was motivated by two facts: 1) several examples from Middle English did not conform to current theories of the pragmatics of object topicalization for Modern English; and 2) at the same time that topicalization was declining,

---

[2]I have updated some of the spelling to make this example easier to read.

English was losing subject-verb inversion (see (1) and (2) above). These two diachronic facts could be related as Yiddish, a V2 language, has much freer constraints on object fronting (see Prince, 1999 for a discussion). As English lost V2, it could have lost such unconstrained topicalization, perhaps as a result of some kind of pragmatic requirement that the object be pre-verbal while the subject post-verbal, no longer satisfied by the OSV order of the non-V2 grammar[3]. This change could have manifested itself in an increasing requirement that the preposed object be in a poset relationship with other entities in the discourse model. However, the decline in topicalization appears to have occurred in all discourse environments equally. This is evidenced by the stasis in the relative distribution of the topic types examined.

Tokens of topicalization from Modern English corpora do not always have a preposed entity that is in a clear poset relationship with other entities in the discourse, nor do they always involve a salient OP. These cases provoke the question: Is there a theory of topicalization that can account for naturally-occurring examples of topicalization that do not mark an OP as salient or those with topics in functional dependence relations (and other non-poset, non-identity type preposings) while precluding topicalizations that are infelicitous? A more extensive study of modern data, preferably with access to intonation, will be necessary to explore this issue.

References

Birner, Betty J. and Gregory Ward. 1998. *Information Status and Noncanonical Word Order in English.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

British National Corpus. The BNC data herein has been extracted from the British National Corpus Online service, managed by Oxford University Computing Services on behalf of the BNC Consortium. All rights in the texts cited are reserved.

Francis, W.N. and H. Kucera. A Standard Corpus of Present-Day Edited American English, for use with Digital Computers. Brown University, Providence, Rhode Island. Department of Linguistics, Brown University 1964, Revised 1971, Revised and Amplified 1979, tagged form 1982. Distributed by the International Computer Archive of Modern English (ICAME).

Hirschberg, Julia. 1991. *A Theory of Scalar Implicature.* New York: Garland.

---

[3] An in-depth analysis of the relationship between the loss of V2 and the decline in topicalization is presented in Whitton (2003).

Johnson, Daniel and Laura Whitton. 2002. A Corpus-Based Quantitative Analysis of Subject-Verb Inversion in Middle and Early Modern English. Ms., University of Pennsylvania.

Kroch, Anthony and Ann Taylor, Eds. 2000. Penn-Helsinki Parsed Corpus of Middle English. (Second edition.) Philadelphia: Department of Linguistics, University of Pennsylvania.

Kroch, Anthony and Beatrice Santorini. (in preparation). Penn-Helsinki Parsed Corpus of Early Modern English. Philadelphia: Department of Linguistics, University of Pennsylvania.

Prince, Ellen. 1999. How not to mark topics: 'Topicalization' in English and Yiddish. *Texas Linguistics Forum*. Austin: University of Texas.

Prince, Ellen. 1997. On the functions of Left-Dislocation in English discourse. In Kamio, A., ed. *Directions in Functional Linguistics,* 117–44. Philadelphia/ Amsterdam: John Benjamins.

Prince, Ellen. 1986. On the syntactic marking of presupposed open propositions. In *Papers from the Parasession on Pragmatics and Grammatical Theory, 22nd Regional Meeting, Chicago Linguistics Society.,* Ed., A. Farley, P. Farley, and K.E. McCullough. 208–222

Speyer, Augustin. To appear. Topicalization and the Trochaic Requirement. *Penn Working Papers in Linguistics*, 10.2.

Vallduví, Enric, and Maria Vilkuna. 1998. On Rheme and Kontrast. In Culicover, P. and McNally, L., Eds., *Syntax and Semantics. The Limits of Syntax.* 29. NY: Academic Press, 79–108.

Webber, B. 1998. Discourse Deixis: Reference to discourse segments. In *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*, SUNY Buffalo, Buffalo NY, June 1988, 113–122.

Whitton, Laura, 2003. An Examination of the Relationship between the Pragmatics of Preposed Objects and the Decline in Both Topicalization and Subject-Verb Inversion in the History of English. Ms., University of Pennsylvania.

Department of Linguistics
University of Pennsylvania
Philadelphia, PA 19104-6305
lwhitton@ldc.upenn.edu