

Unexpected left dislocation: An English Corpus Study

Emily Manetta

Asian Studies Program, University of Vermont, 479 Main Street, Burlington, VT 05405, United States

Received 26 October 2005; received in revised form 13 October 2006; accepted 7 January 2007

Abstract

The discourse function of the syntactic construction left dislocation in English has received significant attention. Prior research has identified at least three distinct form-function correlations underlying left dislocation. This paper examines left dislocation tokens from a corpus of spoken English recorded in South Philadelphia. From this emerges a fourth type of left dislocation not previously identified. We define this variety of left dislocation, termed the Unexpected Subject type, via a Centering Theory analysis of the surrounding discourse. This finding adds even greater diversity to the potential discourse functions underlying the left dislocation construction, and thus lends important support to the claim that the association between syntactic form and discourse function is arbitrary. Future research utilizing much larger corpora will not only allow us to hone the definition of the Unexpected Subject type, but may also reveal that there are indeed more discourse functions of the left dislocation construction that have not been identified.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Form-function correlation; Left dislocation; Discourse function; Centering Theory; South Philadelphia English

1. The discourse function of left dislocation

The discourse function underlying the left dislocation construction in English has received significant attention (Ross, 1967; Gundel, 1985; Geluykens, 1992). A left dislocated NP is one appearing in an initial pre-clausal position that is coreferential with a personal pronoun that occurs elsewhere in the clause. The left dislocated phrases are bolded in (1)–(3).

- (1) **My turtle**, he froze to death. (CL, 10/14/00)¹
- (2) **My aunt Vicky**, she used to get bird poop on her every year. (BB 10/14/00)
- (3) **The knife**, I kept it at school. (BD, 12/6/00)

E-mail address: emily.manetta@uvm.edu.

¹ References to the South Philadelphia Corpus will be made in this format, with the initials of the speaker's pseudonym followed by the date the utterance was recorded.

Left dislocation, probably first identified by Ross (1967), is typically understood as a conversational syntactic construction that has a particular discourse function (or more often a set of discourse functions). Frequently mentioned as foremost among these functions is the introduction of referents that are in some way new to the discourse. Prince's (1997, 1998) approach to left dislocation is situated in a wider claim that discourse function cannot simply be 'read off' syntactic form. Instead, this account asserts that the relation between the two is entirely arbitrary and language-specific. In support of this view, Prince examines a range of left dislocation data in order to refute the notion that left dislocation in English has a single 'contrasting' or 'topic-marking' function. Prince clearly identifies three form-function correlations, described below.

The first type of left dislocation identified by Prince is termed Simplifying Left Dislocation. Consider the token in (4).

- (4) We went to Florida last summer, and we went to Disney World. The best ride the whole time was Jurassic Park. It was so scary. **My sister Chrissie**, her eyes were poppin' out. (CL 10/14/00)

In (4), the bolded material *my sister Chrissie* has not yet been mentioned in the discourse string, and is therefore discourse-new. Were the sentence in canonical form, the left dislocated material would have been in subject position: *my sister Chrissie's eyes were poppin' out*. Prince makes use of Gundel's (1985) and Reinhart's (1981) topic tests and determines that this type of left dislocation does not indicate topichood. Prince's definition of Simplifying Left Dislocation is in (5).

- (5) Simplifying Left Dislocation
A 'Simplifying' Left Dislocation serves to simplify the discourse processing of discourse-new entities by removing the NPs evoking them from a syntactic position disfavored for NPs evoking discourse-new entities and creating a separate processing unit for them. (Prince, 1998:6)

A second function of left dislocation, according to Prince, is entirely distinct from the simplifying role. Let us consider the token in (6).

- (6) So I see my youngest brother a lot too. Actually, all my brothers are pretty close by. My oldest brother is a chef, like, downtown and my middle brother lives in Jersey. **My youngest brother**, he's a freshman at Newnan. (GS, 11/00)

Immediately the contrast between (4) and (6) is clear: *my youngest brother* is not discourse-new. In fact, it is mentioned by the speaker several sentences before the left dislocation. In (4), the referent of the left dislocated constituent *my youngest brother* is actually a member of a set of entities mentioned in the sentence, including *my oldest brother* and *my middle brother*. According to Prince, the left dislocation of *my youngest brother* has the function of indicating the relationship of this entity to other entities in the discourse. This is stated more formally in (7).

- (7) Poset Left Dislocations
A Poset Left Dislocation triggers an inference on the part of the hearer that the left dislocated NP represents an entity standing in a salient partially-ordered set relation to entities already evoked in the discourse. (Prince, 1998:6)

The third role of left dislocation as identified by Prince is actually topicalization in another guise, or Island Amnesty (Resumptive Pronoun Topicalization). According to Prince, these types of left dislocations occur when topicalization would have been appropriate, except that the extraction site is ungrammatical. A resumptive pronoun thus salvages these instances of topicalization, and makes them appear as left dislocation. Prince cites the following example from Studs Terkel:

- (8) **My first book**, I paid half of each trick to the person who gave it to me.
After that, it was my book.’ (Terkel, 1974:95)

In (8), the site of extraction would have been inside a relative clause, and so topicalization would have been impossible. Since this function of left dislocation is less relevant to this paper, I will simply restate Prince’s characterization that this particular type of left dislocation occurs when: “Topicalization is warranted on discourse grounds but is difficult or impossible on grammatical grounds and where a resumptive pronoun occurs in the extraction site, giving the illusion of a Left Dislocation” (Prince, 1998:15).

2. A new type of left dislocation in the South Philadelphia Corpus

The South Philadelphia Corpus (Manetta et al., 2001) consists of approximately 250 min of spontaneous speech, recorded in 2000–2001. The data in the corpus was primarily produced through the use of the face-to-face sociolinguistic interview in which linguistically naïve subjects are asked a range of questions about their neighborhood, background, and important personal experiences.

The speakers in the corpus are themselves all natives of South Philadelphia, inhabiting a four-block section of the tight-knit Italian-American neighborhood. The subset of the corpus used for this paper consists of individual interviews with three adult males, ages 19, 21, and 46, as well as group interviews with eight children of both genders ranging in age from 8 to 13.

There were a significant number of left dislocations (61) in this segment of the South Philadelphia corpus, given its size. Table 1 shows a breakdown of the left dislocations by two of the three types identified by Prince (1998): Simplifying, and Poset. No tokens of the Amnesty type were identified.

As Table 1 shows, the majority of the instances of left dislocation were of the simplifying type (62.9%), such as the token in (4). A smaller, but appreciable number of Poset Left Dislocations were also identified (30.6%), including the token in (6). Of greatest interest to this paper are the four instances, discussed in detail below, which did not appear to fall under either the Simplifying or Poset types as defined by Prince (nor were they Amnesty examples).

Table 2 breaks down the left dislocations by speaker and further by left dislocation type. Initials of each speaker’s pseudonym are given for the adult speakers, while the abbreviation *KIDS* stands for all the children in the group interviews.

Table 1
Left dislocation by type in the South Philadelphia Corpus

Type	Number
Simplifying	39
Poset	19
Other	4
Total	62

Table 2
Left dislocation by speaker and type in the South Philadelphia Corpus

Speaker	Number	Simplifying	Poset	Other
BD	14	9	4	1
GS	8	2	5	1
JC	11	4	7	0
KIDS	29	24	3	2

As Table 1 indicates, there were four left dislocations in the corpus that did not conform to any of the definitions of the three left dislocation types identified by Prince (1998). Those four tokens are listed below in (9)–(12):

- (9) Speaker A: Tell him about Bingo. Tell him about your iguana.
Speaker B: **My iguana Bingo**, he almost bit my finger off. (CL, BB 10/14/00)
- (10) Speaker A: He broke his leg and he had a cast up to here.
Speaker B: **My cast**, it had signatures all over. (CL, GL 10/14/00)
- (11) He didn't need the money . . . He said, I want you to buy it 'cause I know you'll keep it open. My dad talked to the guy. And **the guy who owned it**, he got a loan. (GS, 11/9/00)
- (12) I called him on the phone, to like meet him for the first time. Like, he lives in New Hampshire. He went to boarding school. **Boarding school**, it's where like big rich kids – [cutoff] (BD, 12/6/00)

The left-dislocated NPs are bolded in (9)–(12). None appear to be of the Simplifying type, as defined in (5). Although the personal pronoun coreferential with all of these NPs is in subject position (that is, these NPs would be in subject position were the sentence in canonical order), none of these NPs can strictly be called discourse-new. In the case of the token in (9) the proper name *Bingo* is mentioned in Speaker A's utterance, followed by a mention of *iguana*, so that the entire NP left dislocated by Speaker B is discourse-old. In (11), the referent denoted by *the guy who owned it* is identified using the pronoun *he* in the first and second sentences in the discourse string reproduced here, then as *the guy* in the sentence immediately preceding the left dislocation. (10) and (12) have similar characteristics.

At the same time, none of the left-dislocated NPs in (9)–(12) are in a salient partially-ordered set relation to other entities evoked in the discourse. For instance, *my iguana Bingo* is not in a poset relation to other pets, other iguanas, or even other Bingos. In fact, the discourse in this case is concerned with dangerous things that happened to the speakers. The left dislocations in (10)–(12) are much the same in this respect. Finally, since none of these left dislocations represent environments in which topicalization would have been appropriate but was impossible, as defined by Prince, these are also not left dislocations of the Island Amnestying type.²

At this point we must agree that the left dislocations in (9)–(12) are of an as yet unidentified type, and require closer examination. However, in what follows I will choose to disregard the tokens in (9)–(10). Because in each case the left dislocation is uttered at the outset of a new

² Since, as discussed in the paragraph above, these left dislocated NPs do not stand in any poset relation to entities already evoked in the discourse-model, this cannot be a failed topicalization environment.

speaker turn, it could be claimed that the left-dislocated NP is in fact discourse-new in that specific discourse segment (see Keenan and Shieffelin, 1976). If this were the case, it could potentially be viewed as a left dislocation of the Simplifying type. I will focus instead on defining and identifying the characteristics of the left dislocations in (11)–(12).

It is important to note here that there is no reason to believe that this new type of left dislocation is limited to the variety of English recorded in the South Philadelphia corpus. Geluykens (1992) groups left dislocations that do not appear to be referent-introducing into an “Other” category, such as the token in (13).

- (13) **This guy**, if he did go through three filing cabinets looking for where the petty cash was, he was either extraordinarily neat and taking his time, or else he didn’t bother. (Geluykens, 1992:95)

Geluykens informs us that the left dislocated NP *this guy* in (13) is not discourse-new. It is not being contrasted to any other entity, or in a poset relation with a set of entities. Also, the coreferent pronoun is in subject position. While it is difficult to be certain without access to the entire discourse string, this token is possibly a left dislocation of this new type as well.

3. A centering analysis of left dislocation

Let us now take a closer look at the discourse surrounding the left dislocations in (11)–(12), repeated here. The left dislocated NP has been bolded, and all previous NPs or pronouns referring to the same entity as the left dislocated NP are italicized.

- (11) *He* didn’t need the money . . . *He* said, *I* want you to buy it ‘cause *I* know you’ll keep it open. My dad talked to *the guy*. And **the guy who owned it**, he got a loan. (GS, 11/9/00)
- (12) I called him on the phone, to like meet him for the first time. Like, he lives in New Hampshire. He went to *boarding school*. **Boarding school**, it’s where like big rich kids – [cutoff] (BD, 12/6/00)

Because these two tokens of left dislocation do not conform to any of the three types identified by Prince, we need to find a way to identify the discourse function employed here. In this section, I will show that Centering Theory (Grosz et al., 1995) provides a way to define this fourth type of left dislocation and understand it in relationship to the types already identified, by offering insight into the way each utterance in the discourse relates to adjacent utterances.

According to Centering Theory, each utterance has centers linking it to the surrounding discourse. For every utterance there is a set of forward looking centers $C_f(U)$, and there is a backward looking center $C_b(U)$. Here I will use C_p to indicate the most highly ranked forward looking center. The forward looking center of an utterance is entirely dependent on the expressions in that utterance. An entity in the subject position in an utterance is likely to be the C_p of that utterance (Grosz et al., 1995).

The backward looking center of an utterance U_{n+1} is the most highly ranked element of $C_f(U_n)$ that is realized in U_{n+1} . The backward looking center of an utterance connects in an important way with the forward looking center of the previous utterance. Grosz, Joshi, and Weinstein define three different types of transitions between utterances:

- (14) CONTINUE— $C_b(U_{n+1}) = C_b(U_n)$ and this entity is the $C_p(U_{n+1})$
 RETAIN— $C_b(U_{n+1}) = C_b(U_n)$ and this entity is not the $C_p(U_{n+1})$
 SHIFT— $C_b(U_{n+1}) \neq C_b(U_n)$

A CONTINUE is characteristic of the smoothest discourse, and is more desirable than a RETAIN transition, which is in turn more desirable than a SHIFT. We can now explore a centering analysis of the interesting left dislocation tokens.

The sentence-by-sentence breakdown in (15) presents a centering analysis of the discourse in (11).

- (15) a. He said, I want you to buy it. [want segment] $C_b = \text{the guy}/C_p = \text{the guy}$
 b. My dad talked to the guy. $C_b = \text{the guy}/C_p = \text{dad}$ [RETAIN]
 c. And the guy who owned it, he got a loan. $C_b = \text{the guy}/C_p = \text{the guy}$

In the transition between (15a) and (15b) there is a RETAIN. This means that the hearer expects the topic to change from *the guy* to *dad* as the discourse continues. However, the utterance which follows flouts this expectation—in it both $C_b = \text{the guy}$ and $C_p = \text{the guy}$. That is, the hearer does not expect that the subject of (15c) will be *the guy*.

The sentence-by-sentence breakdown in (16) is the centering analysis of the discourse in (12).

- (16) a. He lives in New Hampshire. $C_b = \text{friend}/C_p = \text{friend}$
 b. He went to boarding school. $C_b = \text{friend}/C_p = \text{friend}$ [CONTINUE]
 c. Boarding school, it's where like big rich kids— $C_b = \text{boarding school}$

The transition from (16a) to (16b) is not a RETAIN, but instead a CONTINUE. That is, it is the hearer's expectation that the speaker will go on talking about the *friend*. (16c) flouts this expectation—the speaker SHIFTS to *boarding school*. Again, the subject *boarding school* is not expected by the hearer in (16c).

In both (11) and (12), the left dislocated NPs are in some sense unexpected as subjects of their respective sentences.³ That is, the hearer is unprepared for these subjects due to the structure of the preceding discourse. I will term this type of left dislocation the 'Unexpected Subject' type, and will define it as follows:

- (17) Unexpected Subject Left Dislocation
 An Unexpected Subject Left Dislocation serves to simplify the discourse processing of entities which the hearer does not expect in subject position due to the structure of the preceding discourse by removing these NPs and creating a separate processing unit for them.

In left dislocations of the Unexpected Subject type, the discourse function of the left dislocation is very similar to that of the Simplifying type.⁴ Simplifying Left Dislocations serve to remove a discourse-new NP from a position disfavored for such NPs. Unexpected Subject Left Dislocations

³ It may be possible to formulate a Centering definition of the Unexpected Subject type. Perhaps $C_p(U_n) \neq C_b(LDU_{n+1})$ would suffice, although many more tokens would need to be examined.

⁴ A reviewer has pointed out that we could potentially state the Simplifying discourse function in Centering terms such that it includes the Unexpected Subject function (many Simplifying Left Dislocations seem to follow a SHIFT). Again, a large number of tokens would need to be examined to test any such definition. We should also note that a centering definition alone may not allow us to distinguish between discourse-new and discourse-old (but unexpected) dislocated constituents.

serve to remove an NP which would otherwise be an Unexpected Subject in the discourse from that position. According to Prince's logic, in both cases the left dislocation construction serves to create a separate unit from the problematic sentence in which the NP can first be processed.

Given the addition of left dislocations like those in (11)–(12), we can now designate at least four known discourse functions that correspond to this syntactic form.

- (18)
- i. Simplifying Left Dislocations
 - ii. Unexpected Subject Left Dislocations
 - iii. Poset Left Dislocations
 - iv. Island Amnesty: Resumptive Pronoun Topicalization Left Dislocations

The implications of this finding are largely concerned with Prince's (1998) core claim: that the connection between syntactic form and discourse function are language-specific and arbitrary. Through her original classification of left dislocation constructions into three types, Prince attempted to show that the syntactic construction left dislocation plays no single discourse-related role. The addition of a fourth type of left dislocation in English serves to lend further credence to this claim—left dislocation seems to serve a wide variety of discourse functions and is motivated by a range of discourse circumstances.

Acknowledgments

I would like to thank Daniel Johnson and Emily Stetler, as well as the other members of the Study of the Speech Community research group at the University of Pennsylvania in 2000–2001. Thanks are also due to our informants, who gave their time and shared their experiences with us. Previous versions of this work were presented at the Pragmatics Workshop at University of Pennsylvania and in the Centering Theory Reading Group at the University of California, Santa Cruz, and I have benefited from the comments of these audiences. Of course, all remaining errors are my own.

References

- Geluykens, Ronald, 1992. From Discourse Process to Grammatical Construction: On Left Dislocation in English. John Benjamins, Amsterdam.
- Grosz, Barbara, Joshi, Aravind, Weinstein, Scott, 1995. Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics* 21 (2), 203–225.
- Gundel, Jeanette, 1985. 'Shared knowledge' and topicality. *Journal of Pragmatics* 9, 83–107.
- Keenan, Elinor, Shieffelin, Bambi, 1976. Foregrounding referents: a reconsideration of left-dislocation in discourse. In: Thompson, H., et al. (Eds.), *Proceedings of the Second Annual Meeting of the Berkeley Linguistics Society*, Berkeley, CA, pp. 240–257.
- Manetta, Emily, Johnson, Daniel, Stetler, Emily, 2001. South Philadelphia Corpus.
- Prince, Ellen, 1997. On the functions of left-dislocation in English discourse. In: Kamio, A. (Ed.), *Directions in Functional Linguistics*. John Benjamins, Philadelphia, pp. 117–144.
- Prince, Ellen, 1998. On the limits of syntax, with reference to topicalization and left-dislocation. In: Cullicover, P., McNally, L. (Eds.), *Syntax and Semantics*, vol. 29. Academic Press, New York, pp. 281–302.
- Reinhart, Tanya, 1981. Pragmatics and linguistics: an analysis of sentence topics. *Philosophica* 27 (1), 53–94.
- Ross, John R., 1967. Constraints on variables in syntax. MIT Ph.D. Dissertation.
- Terkel, Studs, 1974. *Working*. Avon, New York.

Emily Manetta has recently completed her PhD at the University of California, Santa Cruz. Her primary focus is in formal syntax of lesser-studied languages of the Indic group, with a secondary interest in field research on varieties of American English. She currently teaches at the University of Vermont.