Using a probabilistic model of discourse relations to investigate word order variation

Cassandre Creswell
Cymfony, Inc.
600 Essjay Rd
Williamsville NY 14221
USA
creswell@ling.upenn.edu

Abstract
Like speakers of any natural language, speakers of English potentially have many different word orders in which to encode a single meaning. One key factor in speakers’ use of certain non-canonical word orders in English is their ability to contribute information about syntactic and semantic discourse relations. Explicit annotation of discourse relations is a difficult and subjective task. In order to measure the correlations between different word orders and various discourse relations, this project utilizes a model in which discourse relations are approximated using a set of lower-level linguistic features, which are more easily and reliably annotated than discourse relations themselves. The featural model provides statistical evidence for the claim that speakers use non-canonicals to communicate information about discourse structure.

1 Introduction: Non-canonical main clause word order in English

Users of natural languages have many ways to encode the same propositional content within a single clause. In English, besides the “canonical” word order, (1), options for realizing a proposition like GROW(MYRA, EGGPLANTS), include topicalization, left-dislocation, it-clefts, and wh-clefts, shown in (2–5), respectively.

(1) Myra grows eggplants.
(2) Eggplants, Myra grows.
(3) Eggplants, Myra grows them.
(4) It’s eggplants that Myra grows.
(5) What Myra grows are eggplants.

Corpus-based research has shown that these forms are appropriate only under certain discourse conditions (Prince, 1978; Birner and Ward, 1998); among others. These include the membership of referents in a salient set of entities (left-dislocations and topicalizations) or the salience of particular propositions (topicalizations and clefts). For example, in (6), the topicalization is felicitous because there is a salient set KINDS OF VEGETABLES and a salient open proposition, that Myra stands in some relation X with an element of that set.

(6) Myra likes most vegetables, but eggplants she adores.

V = {KINDS OF VEGETABLES};
P = X(m_1, v_2), SUCH THAT v_2 ∈ V

The discourse conditions licensing the use of these non-canonical syntactic forms are necessary conditions. When they do not hold, native speakers judge the use of the form infelicitous. They are not, however, sufficient conditions for use because salient sets and open propositions are ubiquitous in any discourse context, but these non-canonical forms are rare. Each type alone makes up < 1% of utterances, across a variety of genres (Creswell, 2003).

In addition to their information structure functions, one additional communicative goal these word orders fulfill is providing information about how an utterance is related to other discourse segments (Creswell, 2003). Native speaker intuitions about the appropriateness of non-canonicals in particular contexts provide anecdotal evidence (i.e. based on listing individual examples) for this discourse function. To provide broader support for this claim, however, we need to be able to generalize across many tokens.

Ideally, a corpus annotated with discourse relations would be used to measure the correlations between the presence of non-canonical word order and particular discourse relations. However, explicit annotation of discourse relations is a difficult task, and one heavily dependent on the specific theory from which the set of discourse relations is chosen. Instead, this paper describes how a set of more easily-annotated features can be used to create a simplified approximation of the discourse context surrounding non-canonical (or canonical control) utterances. These features are then used as the indepen-
dent variables in a statistical model which provides evidence for claims about how speakers use non-canonical word order to communicate information about discourse relations.

The remainder of the paper is organized as follows: Section 2 describes how some non-canonical word orders in English contribute to the establishment of certain discourse relations. Section 3 describes how these relations can be approximated with a probabilistic model composed of more easily annotated features of the discourse context. Section 4 presents results and discussion of using such a model to measure the correlations between discourse relations and word order. Section 5 concludes and suggests improvements and applications of the model.

2 Additional meaning of non-canonical syntax: discourse relations

The meaning of a multi-utterance text is composed not only of the meaning of each individual utterance but also of the relations holding between the utterances. These relations have syntactic aspects, such that single utterances can be grouped together and combined into segments recursively and are often modeled as a hierarchical tree structure (Grosz and Sidner, 1986; Webber et al., 1999). Discourse relations may also have a semantic or meaning component; this property, when treated in the literature, is often referred to as coherence, subject matter, or rhetorical relations (Kehler, 2002; Halliday, 1985; Mann and Thompson, 1988).

The use of an utterance with non-canonical word order helps hearers make inferences about both the syntactic and semantic properties of discourse relations between the utterance and the rest of the discourse. For both aspects of discourse relations, it is the fact that the non-canonical order marks part of the utterance’s information as salient or discourse-old that assists these inferences.

2.1 Syntax of discourse relations

One substructure of a coherent discourse structure is its attentional structure, which can be modeled as a stack of focus spaces (Grosz and Sidner, 1986). Each segment in the discourse tree has a corresponding focus space containing the currently salient discourse entities. When a segment begins, its focus space is pushed onto the stack on top of any other incomplete segments’ spaces. When the segment ends, the focus space is popped off the stack. When an utterance continues in the same segment, the focus stack is unchanged.

Non-canonical utterances instruct hearers about where to attach segments to the discourse tree. Because of the necessary conditions that license the use of a non-canonical, in most cases the open proposition or set is part of a focus space pushed onto the stack previously. So, the non-canonical form evokes the old proposition or set and thus reactivates the salience of that focus space. Reactivating the salience of the focus space in turn activates the salience of the discourse segment. As a result, the hearer infers that the new segment associated with the non-canonical utterance should be attached at the same level as this reactivated discourse segment, i.e. at a non-terminal node on the tree’s right frontier. Any intervening segments should be closed off, and their focus spaces should be popped off the stack.

To illustrate, in (7) the use of the it-cleft occurs after an intervening discussion of a separate topic. It-clefts are used to indicate that an existential closure of an open proposition is presupposed, here ∃!YOU GOT TO MICHIGAN STATE AT TIME t. This presupposed material allows speaker B to mark the question as related to the prior discussion. In a tree structure of this discourse, the cleft corresponds to an instruction to “pop” back to a higher level in the tree when attaching the utterance, where speaker G’s career at Michigan State was under discussion. The canonical version in (8) is an abrupt and infelicitous continuation of the discourse, as if B is unaware of the previous discussion of G’s arrival at Michigan State.¹

(7) G: So for two years, I served as a project officer for grants and contracts in health economics that that agency was funding. I decided to go to academia after that and taught at Michigan State in economics and community medicine. One thing I should mention is that for my last three months in government, I had been detailed to work on the Price Commission which was a component of the Economic Stabilization program. [Description of work on Price Commission...]

B: In what year was it that you got to Michigan State? (SSA, ginsberg)

(8) In what year did you get to Michigan State?

2.2 Semantics of discourse relations

The contribution of non-canonical utterances to the inference of semantic aspects of discourse relations is also related to the fact that these word orders mark (part of) an utterance’s content as discourse-old or presupposed. Non-canonical word order is

¹Varying the placement of the primary prosodic stress may improve the version in (8); see Delin (1995) and Creswell (2003) for comparison of the discourse function of prosody and syntax.
used to indicate relations of RESEMBLANCE rather than CONTIGUITY.

A CONTIGUITY relation is the basic relation found in narratives. According to Labov (1997), utterances in a prototypical narrative describe in the order they took place a sequence of causally-related events which lead up to a MOST REPORTABLE EVENT. Kehler (2002), following Hobbs (1990), says the events should be centered around a system of entities, and each event should correspond to a change of state for that system. To infer a CONTIGUITY relation between two utterances, the hearer must infer that their eventualities correspond to a change of state for that system.

Inferring a RESEMBLANCE relation between two utterances depends on a very different type of information. To establish RESEMBLANCE, the hearer must identify a common relation $R$ that relates the propositional content of two utterances and also the number and identity of their arguments (Kehler, 2002). Relevance relations include PARALLEL-, CONTRAST, EXEMPLIFICATION, GENERALIZATION, EXCEPTION, and ELABORATION.

Non-canonicals are useful in resemblance relations because 1) the presence of ‘old’ material in a non-canonical helps overrule the default coherence relation of CONTIGUITY by making that interpretation less likely, and 2) the use of old material and a structured proposition assists the hearer in identifying a common relation and corresponding arguments needed to establish RESEMBLANCE.

This is illustrated in (9). The use of a left-dislocation tells the hearer that the referent of a lot of the doctors is in a salient set. By identifying that set as $\{\text{PROFESSIONAL PEOPLE}\}$, the hearer can realize that the information being added about a lot of the doctors is going to be in an EXEMPLIFICATION relation with the earlier statement that professional people in general began to think of themselves as disabled.

(9) During the Depression an awful lot of people began to think of themselves as disabled, especially professional people, who depended on clients whose business was on a cash basis—there was no credit, this was a universe without credit cards. A lot of the doctors, they were doing an awful lot of charity work. They couldn’t support themselves. They’d have a little heart attack. They’d have disability insurance. They went on the insurance company rolls. A lot of doctors had disability insurance and a lot of others too. A lot of the insurance companies stopped underwriting disability insurance. They couldn’t afford it. (SSA, hboral)

(10) A lot of the doctors were doing an awful lot of charity work.

Without the left-dislocation, identifying the inclusion relationship between the set of professional people and doctors is quite difficult. The preferred interpretation of the canonical version in (10) is only that the doctors were doing charity work for professional people who had no credit cards. The left-dislocation supports the additional inference that the exemplification described above holds too.

3 Probabilistic model of discourse relations and non-canonical syntax

To provide evidence beyond individual examples for the phenomena in Section 2, we need to measure the correlation between discourse relations and syntactic form, but annotating discourse relations directly is problematic. Annotation of hierarchical discourse structure is difficult and subjective although efforts have been made (Creswel et al., 2002; Marcu et al., 1999). Even annotating linear segmentation is challenging, particularly in the vicinity of segment boundaries (Passonneau and Litman, 1997). Annotation of the semantics of discourse relations requires a predetermined set of relation types, on which theories vary widely, making theory-neutral generalizations about the role of non-canonical syntax impossible.

This project attempts to overcome these difficulties by indirectly deriving discourse relations by mapping from their known correlates to the use of certain non-canonical forms. The correlates used here are referential relations across utterance boundaries and the presence and type of lexical discourse markers or cue words. These features are annotated with respect to a three-utterance window centered on a target utterance $U_i$, shown schematically in Figure 1.

These referential and lexical features build on the work of Passonneau and Litman (1997), who use them in discourse segmentation. Their use here is extended to also derive information about the se-
Table 1: Values of referential relations feature

| Ref | Utterances share center of attention; $C_p$ of first utterance is $C_b$ of second utterance. Mary's a vegetarian. She never eats meat. |
|-----|----------------------------------------------------------------------------------------------------------------------------------|
| 3   | Utterances have coreferential NPs.                                                                                             |
| 2   | Mary likes Fred. He's very friendly.                                                                                          |
| 1   | Utterances have inferentially-related NPs.                                                                                  |
|     | *I bought an old bike. The tire was flat. Ostriches are huge. Finches are little.*                                         |
| 0   | Utterances have no NPs that share reference.                                                                                 |

The lexical discourse marker feature is annotated for the target $U_i$ and its preceding $(U_{i-1})$ and following $(U_{i+1})$ utterances and has five values: and, but, so, other or none. The predictions about the correlations between these lexical markers and discourse relations. First, if non-canonicals are indicators of attentional stack pops, they should be more likely at segment boundaries; hence, we expect an increased presence of cue words (Passonneau and Litman, 1997) on non-canonicals compared to canonicals.

Predictions about the type of cue words are based on the survey of lexical cue meanings from Hirschberg and Litman (1994). Because and is an indicator of segment-continuation and the relation CONTIGUITY, we expect decreased incidence on non-canonicals. However, we expect greater incidence on $U_{i+1}$ when $U_i$ is non-canonical because $U_i$ should be used to start a new segment. The presence of but indicates a CONTRAST relation. Non-canonicals should have a greater likelihood of being in contrast with either of the utterances surrounding them,

So we expect a greater incidence of but on both $U_i$ and $U_{i+1}$ for non-canonicals than for canonicals. The presence of so can indicate RESTATEMENT or RESULT, so so should appear more often on $U_i$ for wh-clefts, which are often used in ELABORATION relations.

The referential features are four-valued and annotated with respect to pairs of utterances, $(U_{i-1}, U_i)$ and $(U_i, U_{i+1})$. The values here, described in Table 1, form an implicational scale from strongest to weakest connections, and the utterance pair is labeled with the strongest relation that holds.

In order to decide whether the featural models provide evidence to support the claims about discourse relations and syntactic forms, we first need to make clear our assumptions about how referential relations and lexical markers correlate with discourse relations. Based on those assumptions, testable predictions can then be made about how referential relations and lexical markers should correlate with syntactic forms.

2See Creswell (2003) for examples and discussion.
Non-canonicals should be more likely than canonicals to follow a POP and begin a new segment. They should have weaker referential ties to the preceding utterance. They should have a higher incidence of having no referential ties to \( U_{i-1} \), and a lower incidence of having no referential ties to \( U_{i+1} \).

Non-canonicals should be less likely than canonicals to have a NARRATIVE relation with either \( U_{i-1} \) or \( U_{i+1} \). This situation predicts that with respect to both of the utterances surrounding a non-canonical utterance, these utterances will be less likely to share the same center of attention than when \( U_i \) is canonical.

Non-canonicals should be more likely than canonicals to be in RESEMBLANCE relations with \( U_{i-1} \) and/or \( U_{i+1} \). So, a greater likelihood of reference to inferentially-related entities and smaller likelihood of reference to coreferential entities or shared centers in both the preceding and following utterance is expected.

### 4 Results and discussion

To test the predictions about non-canonicals and discourse relations, a corpus of 799 utterances with non-canonical word order were extracted from 58 transcribed interviews from the Social Security Administration Oral History Archives (SSA), a corpus with \( \sim 750,000 \) words and 44,000 sentences. In addition to the four types of non-canonicals, 200 randomly-selected controls with canonical word order were also included. Table 4 lists the breakdown by syntactic type. The two lexical and three referential features described in the previous section were annotated for each of the 999 utterances.

Logistic regression models for binary comparisons between each of the five sentence types were then created. For 9 of 10 comparisons, at least one of the five features were found to be significant.\(^3\)

Table 4 lists all features found to be significant for each of the ten comparisons, i.e. features that individually have a significant effect in improving the likelihood of a model when compared to a model that uses no features to predict the distribution of the two classes.\(^4\) For comparisons with multiple features significant at the five-percent level, the \( p \)-value of the model fit in comparison with a fully saturated model is listed in the fourth column of Table 4.

In order to understand the most likely context in which a form to appears, we need to examine the weights assigned to each feature value by the regression analysis. The detailed feature weights in the best model are listed in Table 5.

Table 6 summarizes the general conclusions we can draw from these weights about the most favorable discourse contexts for each of the four types of non-canonicals. For considerations of space, we discuss in detail only one of the four types here, wh-clefts. Wh-clefts are particularly relevant with respect to the insights they provide into the inherent limitations of our model of discourse relations.

Overall, wh-clefts are favored in contexts where they start a new segment, one with weak connections with the preceding utterance and strong connections with the following utterance. In particular, feature 4, \( \text{Ref}(U_{i-1}, U_i) \), is significant in the

| Syntactic Type | No. of Tokens |
|---------------|---------------|
| It-cleft      | 150           |
| Left-dislocation | 258         |
| Topicalization | 111           |
| Wh-cleft      | 280           |
| Control       | 200           |
| Total         | 999           |

Table 3: Corpus of utterances by syntactic type

\(^3\)The comparison of it-clefts and left-dislocations is the exception here. From the lack of significant features in this comparison we can surmise that the it-clefts and left-dislocations are more similar to each other than any of the other forms compared here.

\(^4\)In particular, the measure whose significance is tested is the \( \Delta -2 \chi^2 \) (difference in log-likelihoods of the models), which is \( \chi^2 \) distributed, where the number of degrees of freedom is the difference in the total number of feature values between the two models.
### Table 2: Predictions from ref. features to discourse relations

| CLASS VS. CLASS | Feature | Feature | Overall Model Fit | p-value |
|-----------------|---------|---------|------------------|---------|
| CONTROL, IT-CLEFT | 2       | 5 (.097) | n.a.             |         |
| CONTROL, LEFT-DIS. | 4, 5    | 3 (.161) | p=0.9289         |         |
| CONTROL, TOPIC.  | 3       | 2 (.178) | n.a.             |         |
| CONTROL, WH-CLEFT | 2, 4    | 3 (.151) | p=0.8696         |         |
| IT-CLEFT, LEFT-DIS. | –       | 1 (.066), 4 (0.086) | –       |         |
| IT-CLEFT, TOPIC.  | 3       | 2 (0.092), 4 (0.099) | n.a.    |         |
| IT-CLEFT, WH-CLEFT | 4       | 1 (0.184) | n.a.             |         |
| LEFT-DIS, TOPIC.  | 3, 4    | 5 (0.129) | p=0.8561         |         |
| LEFT-DIS, WH-CLEFT | 1, 4    | 5 (0.147) | p=0.7615         |         |
| TOPIC, WH-CLEFT  | 2, 3, 4 | –       | p=.6935 (with 3, 4) |         |

Table 4: Features significant at p < 0.05 and p < 0.2. Features significant at p < 0.01 are in **bold**. Features 1, 2, and 3 are discourse marker features on \( U_{i-1}, U_i, \) and \( U_{i+1} \), respectively. Features 4 and 5 are referential features for the pairs \( (U_{i-1}, U_i) \) and \( (U_i, U_{i+1}) \), respectively.

### Table 6: Summary: favorable discourse contexts

| Sentence type | Most Favorable Contexts |
|---------------|-------------------------|
| Topicalizations | CONTINUE with \( U_{i-1} \); CONTRAST with \( U_{i-1} \) or \( U_{i+1} \) |
| Wh-clefts | POP after \( U_{i-1} \); CONTRAST or CONTINUE with \( U_{i+1} \) |
| Left-dislocations | POP after \( U_{i-1} \) or RESEMBLANCE with \( U_{i-1} \); CONTINUE with \( U_{i+1} \) |
| It-clefts | No strong tendencies for begin/end of segments; possible CONTRAST relations with \( U_{i-1}, U_{i+1} \) |

comparison of wh-clefts with all other classes. Wh-clefts are much more likely to share no connections at all with \( U_{i-1} \) and less likely to share only inferential connections when compared with any other class. In comparison with everything but left-dislocations, wh-clefts are also less likely to share their center of attention with \( U_{i-1} \).

In terms of discourse markers, feature 2 and 3 are significant when comparing topicalizations and controls with wh-clefts (although feature 3 is only weakly significant in comparing wh-clefts and controls.) For feature 2, MARKER(\( U_i \)), wh-clefts are less likely than either of the other two to appear with and and more likely to appear with so. For feature 3, however, the presence of and on \( U_{i+1} \) favors wh-clefts over topicalizations and controls.

The most likely context in which to find wh-clefts then is one with no referential connections to the previous utterance and marked with the discourse adverbial, so. When the \( U_{i+1} \) begins with and, assumed to be a marker of continuation of the previous content, wh-clefts are also favored. This pattern resembles most closely the descriptions of a preceding discourse POP and a subsequent discourse CONTINUE or NARRATIVE.

One use of wh-clefts that is not borne out conclusively in the data is its use in ELABORATION relations, as in (11). Kehler (2002) describes elaborations as a case of RESEMBLANCE where the predicative and its arguments are the same, but described from a different perspective or level of detail. The hearer must infer the identity of the event and en-
Table 5: Individual feature weights in best model. Feature weights >0.5 favor the application value (class category) listed first; weights <0.5 favor the second application value. The farther away from 0.5, the stronger the feature value’s effect on the distinction between the two classes.

|    | CONTROL | CONTROL | CONTROL | CONTROL | IT-CLEFT | IT-CLEFT | IT-CLEFT | LEFT-DIS. | LEFT-DIS. | LEFT-DIS. | TOPIC. |
|----|---------|---------|---------|---------|----------|----------|----------|----------|----------|----------|-------|
| 1. MARK (U_{i-1}) | a | b | s | o | n | 0.655 | 0.326 | 0.500 | 0.542 | 0.481 |       |
| 2. MARK (U_{i}) | a | b | s | o | n | 0.548 | 0.399 | 0.249 | 0.689 | 0.628 | 0.665 | 0.604 | 0.172 | 0.483 | 0.630 |
| 3. MARK (U_{i+1}) | a | b | s | o | n | 0.514 | 0.266 | 0.408 | 0.738 | 0.574 | 0.595 | 0.232 | 0.406 | 0.685 | 0.602 | 0.603 | 0.268 | 0.470 | 0.641 | 0.533 | 0.343 | 0.804 | 0.579 | 0.311 | 0.429 |
| 4. REF (U_{i-1}, U_{i}) | a | b | s | o | n | 0.418 | 0.443 | 0.511 | 0.627 | 0.574 | 0.352 | 0.575 | 0.493 | 0.583 | 0.334 | 0.639 | 0.566 | 0.634 | 0.420 | 0.279 | 0.574 | 0.467 | 0.574 |
| 5. REF (U_{i}, U_{i+1}) | a | b | s | o | n | 0.785 | 0.401 | 0.370 | 0.409 | 0.785 | 0.401 | 0.370 | 0.409 | 0.785 | 0.401 | 0.370 | 0.409 | 0.785 | 0.401 | 0.370 | 0.409 | 0.785 | 0.401 | 0.370 | 0.409 |

Anecdotes being described in the two segments. If wh-clefts are associated with ELABORATIONS, then we should see an increased incidence of close referential connections with U_{i-1} and an increased incidence of so, a marker of restatement. In the results, however, we only see evidence for the latter.

(11) S: How did you develop this Resource-Based Relative Value Scale at this point?
H: We basically treated this as a research project because most of us involved realized we had some past failures and we should not over-promise. We should be prepared to face up to the world and say, “We cannot make the theory operational.” So what we did was we continued to accept the theoretical premise, that is the rational and objective price should be based on the cost of the service. Then we asked, “What constitutes the cost of physicians’ services and what are the components of physicians’ work?” (SSA, hsiao)

A possible factor in the absence of evidence here is that wh-clefts are also associated with discourse pops, which increase the likelihood of having no referential connections with the previous utterance. The logistic regression model used here aggregates over all possible discourse relations. So, when two discourse relations that give rise to different lexical and referential patterns are both associated with a single sentence type, the patterns of one may obscure the patterns of the other. A more sophisticated statistical model might take into account dependencies between discourse markers and referential patterns and from them posit hidden states which correspond to different discourse relations. Then based upon these hidden states, the model would predict which sentence type would best fit the context. Such a model would be more true to Figure 2.

Another limitation of the model shown here is that the only indicators in this model of starting a new segment are weak or absent referential relations, presence of a connective like so, and absence of and. These measures will not necessarily distinguish between continuing in the same segment or beginning a new segment which includes recently-mentioned discourse-old entities.
5 Conclusions and potential applications

The statistical model here uses a combination of referential and lexical features annotated for a small window surrounding the target utterance to represent the local discourse context surrounding utterances with non-canonical and canonical word orders. The primary goal was to model the correlations between discourse relations and non-canonical syntax. Due to the difficulties inherent in annotating discourse relations directly, the featural approximation was devised as a practical alternative.

Overall, the method used here yielded some interesting new insights into the contexts that favor the use of four types of non-canonical word order. The complexity of this approach does make it difficult to draw simple conclusions about the relationship between discourse relations and non-canonical syntactic forms. However, the strength of some of the correlations found here merits further investigation. The data also lend support for the idea that some aspects of discourse relations, both syntactic and semantic, can be inferred from combinations of lower-level linguistic features.

An important factor in improving upon the current project is the need for larger amounts of data. The significance of any particular feature is greatly affected by the quantity of data. This was a particular issue for the lexical feature values, where it prevented inclusion of several of the less frequent connectives with better understood discourse structuring properties, like well and now. In addition, more data may also be required in order to support the use of more complex statistical models. Automatic methods of annotating the referential features or the availability of larger corpora marked up with coreferential and inferential relations and with a rich variety of syntactic forms could be used to test more accurately the predictions in Section 3.

The technique used here for approximating discourse relations through more easily annotated features has at least two interesting potential applications. One, given the significant correlation of these features with non-canonical word order variation, the discriminative models trained here could be used as classifiers which could label discourse contexts (feature vectors) with the form best suited to the context for the surface realization stage in a natural language generation system.

Secondly, the feature set used here could be applied to the problem of automatic classification of discourse relations. In conjunction with a relatively small set of pairs of sentences for which there is high inter-annotator agreement when hand-annotated for type of discourse relation, the lexical and referential features here could serve as an initial feature set for bootstrapping the development of a statistical discourse relation classifier. This application would require stipulation of a predetermined set of discourse relations—a requirement the present study wished to avoid. However, given the practical need for a statistical relation classifier, a set of relations could be constructed suitable to the domain of use.

References
Birner, B., and G. Ward. 1998. Information status and non-canonical word order in English. John Benjamins.
Creswell, C. 2003. Syntactic form and discourse function in natural language generation. Doctoral Dissertation, University of Pennsylvania.
Creswell, C., K. Forbes, E. Mitsakaki, R. Prasad, A. Joshi, and B. Webber. 2002. The discourse anaphoric properties of connectives. In Proceedings of DAARC 4, 45–50. Lisbon, Portugal: Edicoes Colibri.
Delin, J. 1995. Presupposition and shared knowledge in it-clefts. Language and Cognitive Processes 10:97–120.
Grosz, B. J., and C. L. Sidner. 1986. Attention, intentions, and the structure of discourse. Computational Linguistics 12:175–204.
Halliday, M. A. K. 1985. An introduction to functional grammar. Baltimore: Edward Arnold Press.
Hirschberg, J., and D. J. Litman. 1994. Empirical studies on the disambiguation of cue phrases. Computational Linguistics 19:501–530.
Hobbs, J. R. 1990. Literature and cognition. Stanford: CSLI.
Kehler, A. 2002. Coherence, reference, and the theory of grammar. CSLI Publishers.
Labov, W. 1997. Some further steps in narrative analysis. Journal of Narrative and Life History.
Mann, W. C., and S. A. Thompson. 1988. Rhetorical Structure Theory: towards a functional theory of text organization. Text 8:243–281.
Marcu, D., E. Amorrortu, and M. Romera. 1999. Experiments in constructing a corpus of discourse trees. In Proceedings of the ACL workshop: Towards standards and tools for discourse tagging, ed. M. Walker, 48–57.
Passonneau, R., and D. Litman. 1997. Discourse segmentation by human and automated means. Computational Linguistics 23:103–139.
Prince, E. F. 1978. A comparison of wh-clefts and it-clefts in discourse. Language 54:883–906.
Webber, B., A. Knott, M. Stone, and A. Joshi. 1999. Discourse relations: a structural and presuppositional account using lexicalised TAG. In ACL 37, 41–48. College Park, MD.