

The Predicate-Argument Structure of Discourse Connectives: A Corpus-based Study

Cassandra Creswell¹, Katherine Forbes², Eleni Miltsakaki¹, Rashmi Prasad¹,
Aravind Joshi¹ and Bonnie Webber³

¹*University of Pennsylvania*, ²*University of Pittsburgh*, ³*University of Edinburgh*

Discourse connectives can be analysed as encoding predicate-argument relations whose arguments derive from the interpretation of discourse units. These arguments can be anaphoric or structural. Although structural arguments can be encoded in a parse tree, anaphoric arguments must be resolved by other means. A study of nine connectives, annotating the location, size, and syntactic type of their arguments, shows connective-specific patterns for each of these features. A preliminary study of inter-annotator consistency shows that it too varies by connective. Results of the corpus study will be used in the development of resolution algorithms for anaphoric connectives.

1 Introduction

Discourse connectives can be analysed as encoding a relation between two discourse segments. In other words, the semantic interpretation of a discourse connective is a predicate that takes discourse units as its arguments. These arguments can be derived anaphorically or structurally. We describe this distinction below in more detail. Roughly, structural arguments can be encoded in a parse tree, but anaphoric arguments must be resolved by other means.

The distinction between anaphoric and structural arguments is a theoretical one based on a discourse lexicalised tree-adjoining grammar (DLTAG). In DLTAG, the compositional part of discourse meaning (projected by the tree structures) is divided from the non-compositional contributions due to general inferencing and anaphora. This division is a key insight of the DLTAG approach to discourse structure which simplifies the set of structures that can be assigned to a discourse.

With respect to any particular connective, its categorization as taking its arguments structurally or anaphorically is an empirical question. Because only structural arguments can be derived from a DLTAG discourse structure, the location of anaphoric arguments is an additional issue that requires empirical investigation of linguistic data. This corpus study is undertaken as a preliminary attempt to annotate discourse connectives' arguments in order to provide evidence for 1) whether the arguments of discourse connectives can be reliably annotated; 2) whether to classify particular connectives as structural or anaphoric; and 3) whether anaphoric arguments of connectives display

properties that would allow development of robust resolution algorithms for locating them.

The results of this corpus study of nine connectives, where the location, size, and syntactic type of their arguments were annotated, sheds light on all three of these issues. First, the data do provide evidence for characterizing certain connectives as anaphoric or structural. In addition, with respect to the features examined here, we found a range of connective-specific behaviours. Finally, a preliminary study of inter-annotator consistency shows that its reliability also varies by connective. The results of this corpus study will be useful for parsing discourse structure, for developing resolution algorithms for anaphoric arguments of connectives, and for revising the annotation guidelines in preparation for a large-scale study of discourse connectives and their arguments.

The structure of the paper is as follows: In Section 2, we provide the theoretical background necessary to understand the distinction of interest here between *structural* and *anaphoric* connectives. This background includes a brief introduction to LTAG and DLTAG. Then, in Section 3, we describe the corpus study undertaken, including its guidelines, results, and an assessment of the reliability of the annotation. In Section 4, we examine the implications that variation in the annotations has for the ability to develop resolution algorithms. We conclude in Section 5 with a discussion of future annotation and algorithm development efforts.

2 Theoretical Background: LTAG and DLTAG

The theoretical background of this study of discourse connectives is Discourse Lexicalised Tree Adjoining Grammar (DLTAG) (Webber *et al.*, 2003). DLTAG builds an intermediate level of discourse structure directly on top of the clause. DLTAG's syntax is currently modelled using the structures and structure-building operations of a lexicalised tree-adjoining grammar (LTAG) (Joshi *et al.*, 1975), which is widely used to model the syntax of sentences.

2.1 LTAG

Briefly, an LTAG is a lexicalised extension of a tree-adjoining grammar (TAG). The object language of an LTAG is a set of trees, allowing the underlying structure of a surface string to be represented, as well as the string itself. An LTAG consists of a finite set of *elementary* trees and operations for combining them. Elementary trees are associated with at least one lexical item, called an *anchor*. They represent extended projections of the anchor and encode its subcategorization frames. An anchor may be associated with more than one tree; each tree in this tree *family* reflects a different syntactic construction in

which that anchor can appear. For example, the verb *eat* may anchor either a transitive or intransitive tree.

There are two types of elementary trees in an LTAG: *initial* trees, which encode basic predicate-argument relations, and *auxiliary* trees, which encode optional modification and must contain a non-terminal node (called the *foot* node) whose label matches the label of the root. The rightmost tree in Figure 1 is an auxiliary tree, all the others are initial trees.

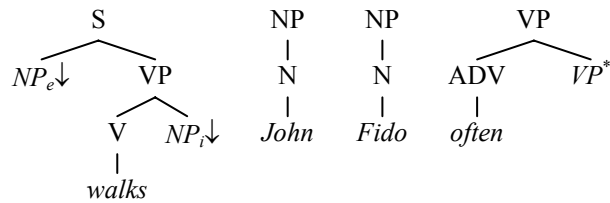


Figure 1: Elementary LTAG trees

There are two structure-building operations in LTAG that create complex trees called *derived* trees: *substitution* and *adjunction*. As shown in Figure 1, substitution sites are indicated by \downarrow and adjunction sites, by $*$.

Substitution consists of replacing the node marked \downarrow with the tree being substituted. Only initial trees or trees derived from initial trees can be substituted, and the root node of the tree being substituted must match the label of the node being replaced. For example, the tree anchored by *Fido* in Figure 1 can substitute for the internal argument (NP_i) in the tree anchored by *walks*, and the tree anchored by *John* can substitute for the external argument (NP_e) in the tree anchored by *walks*. The result of these substitutions is shown in Figure 2(a).

Adjunction is restricted to non-terminal nodes not already marked for substitution, building a new tree from an auxiliary tree β and any other tree τ (initial, auxiliary, or derived). To combine β and τ by adjunction, the root node of β must match the label of the node n in τ to which it is to be adjoined. The root node of β is identified with n ; the subtree dominated by n is attached to the foot node of β , and the rest of the tree that dominated n now dominates the root node of β . For example, the tree anchored by *often* in Figure 1 can adjoin to the VP node in Figure 2(a), producing the derived tree in Figure 2(b).

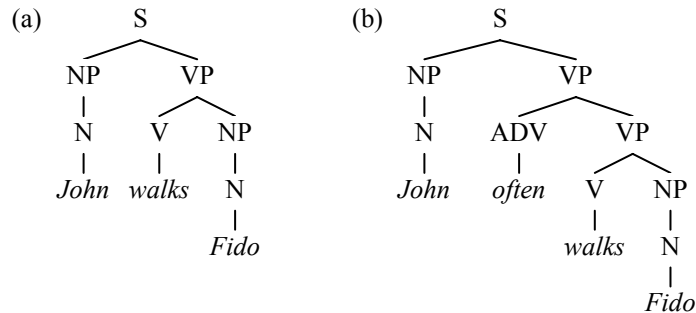


Figure 2: LTAG derived trees after substitution and adjunction

2.2 DLTAG

DLTAG is an extension of LTAG in which the elementary trees are anchored by discourse connectives. Discourse connectives can be analysed as encoding predicate-argument relations whose arguments are the interpretations of discourse segments. The elementary trees anchored by connectives combine with each other and with derived trees to create a structure for a multi-sentence discourse. That is, DLTAG is a grammar for combining sentences into a discourse rather than for combining words into sentences. A lexicalised grammar at the discourse level can capture the inter-sentential relations encoded by connectives and allow an extension of compositional semantic representations from the sentence level to the discourse level.

Just as at the sentential level, arguments to these discourse relations can be found structurally or anaphorically. Here, *structurally* means the semantic content of the argument must be derivable locally. At the sentential level, an example of a relationship with a strictly structural basis is the relationship between a reflexive pronoun and its antecedent, as in (1), where *himself* must co-refer with *John*, the subject of the sentence.

- (1) John saw **himself** in the mirror.

The reflexive pronoun and its antecedent must have a particular relationship to each other in the syntactic tree, one where they are both present in the same clause, i.e. arguments of a single predicate. The antecedent of a free pronoun, however—although there are positions in which it cannot appear with respect to the pronoun—does not need to have any particular local syntactic relationship with the pronoun, and so it can be found within or outside the same sentence. This is illustrated in (2) where *she* may be coreferent with either *Jan* or *Fran*.

- (2) Jan called. Fran said that **she** might come over later.

Locating the antecedent – and therefore computing the interpretation – of a free pronoun relies on anaphoric and inferential mechanisms. This difference

between a structural and an anaphoric relationship at the sentential level is analogous to the one found with the arguments of discourse connectives at the discourse level.

Every discourse connective will find at least one of its arguments structurally, the argument that substitutes into one of the leaf nodes in the tree anchored by the discourse connective. Its other argument may be found either structurally or anaphorically. We will refer to connectives that find one of their arguments anaphorically as *anaphoric connectives*; the others as *structural connectives*.¹ The difference between the two types can be most easily seen in an example where multiple connectives appear together (Webber *et al.*, 2000), like (3).

- (3) S₁: Sally rarely eats meat and subscribes to *Vegetarian Times*.
 S₂: Lately, she's raised the ire of her vegan friends
 S₃: because she nevertheless enjoys the occasional bacon cheeseburger.

In (3), *because*, a structural connective at the discourse level, is the predicate expressing the causal relation between two eventualities, P = RAISE IRE (SALLY, FRIENDS) and Q = ENJOYS (SALLY, CHEESEBURGER). This is encoded formally with the two argument nodes appearing in the same elementary tree, shown in Figure (3).²

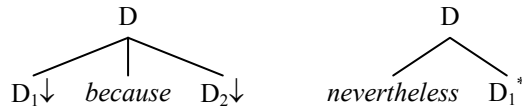


Figure 3: Elementary tree: *because* and *nevertheless*

In contrast, the connective *nevertheless* in S₃ finds only a single argument structurally Q = ENJOYS (SALLY, CHEESEBURGER). Its other ‘left-hand’ argument is derived anaphorically from S₁. The formal way of capturing this difference is assigning a different type of elementary tree to *nevertheless*, also

¹ We view this as a lexical property of a particular connective. If a connective can ever be found with a non-adjacent, non-contiguous, or only inferentially-derivable (rather than textually-derivable) anaphoric argument, it is an anaphoric connective. In a given use of an anaphoric connective, however, its anaphoric argument might occur in the immediately preceding text. As such, to decide what category a particular connective falls into, if a convincing example of an anaphoric use cannot be constructed, then many naturally-occurring examples may also be needed to make this designation.

² In general, the theory has only been applied to monologic text, primarily written rather than spoken. This analysis will have to be extended to account for structural connectives that appear in sentence fragments in dialog, as in *Because I said so*. We suspect that fragments containing structural connectives like *because* will pattern with their structural counterparts in written text rather than with anaphoric connectives, but a detailed study remains for future work.

shown in Figure (3); here, the discourse clause to which the *nevertheless* tree adjoins, D_1 , is its sole structural argument.³

Not all discourse segments (elementary or complex) are related via a lexically explicit discourse connective. In DLTAG, such relations are handled by an auxiliary tree anchored in a lexically-empty discourse connective that conveys *continuation* of the description of the larger tree to which it is attached. Although a more specific relation may be inferred, the relation provided by the syntax alone is semantically underspecified, analogous to the semantics of noun-noun compounds.⁴ In the discourse tree derived from (3), S_2 is attached to the previous discourse with an auxiliary tree anchored by a lexically-empty connective.

The full derived tree for the discourse in (3) is shown in Figure (4).

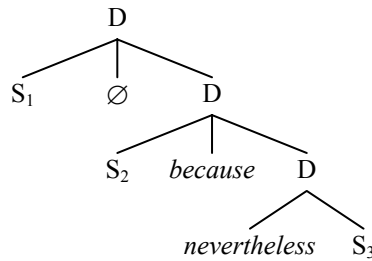


Figure 4: Full discourse tree

We can see that the arguments of structural connectives are encoded directly in a parse tree and, therefore, are relatively easy to identify.⁵ The non-structural argument of an anaphoric connective must be resolved by other means.⁶ Once again, this is similar to the case of bound versus free pronouns.

³ Adverbial connectives may appear sentence initially, medially or finally. The position of the adverbial connective in the sentence affects the scope of the connective and is often associated with the information structure partitioning of the sentence into focus and ground (Kruijff-Korbayová & Webber, 2001). With respect to parsing discourse structure using a DLTAG, in cases of medial and final discourse adverbials, a sentence-initial copy of the adverbial is added during parsing. This makes it possible to use the same elementary tree structure anchored in that lexical connective no matter where the connective appears at the sentence level. An index is retained inside the sentence to retain information about its clause internal position. The discourse-syntactic role then remains the same regardless of its sentence-level syntax. See Forbes *et al.* (2003) for more detail about the use of DLTAG in parsing.

⁴ There may be some limits on the types of relations that may be inferred without the specific use of a discourse connective. Presumably, this will depend on the contributions of sentential semantics, syntax and prosody to the inferential process.

⁵ In fact, structural connectives are associated with attachment ambiguity in the parse tree, and so although once a parse tree is created, identifying them is trivial, the determination of the parse is not itself trivial.

⁶ Although the missing argument will not be in the syntactic tree, it will be represented in the semantics of that tree, e.g. e in the semantics for *nevertheless*:

(1) NEVERTHELESS(x , $[[e_i]]^{ac}$)

Here the semantics links the x argument to an address in the tree, but the e argument is not linked to an address; it is represented using an assignment function. Assignment functions have already been used to

Another property discourse anaphoric connectives share with other types of discourse anaphora is that their anaphoric arguments may be found intra- or inter-sententially, as in (4) and (5), respectively.

- (4) A person who seeks adventure might, for example, try skydiving.
[Webber *et al.* (2000)'s ft.8 (i)]
- (5) Some people seek adventure. For example, they might try skydiving.

Because discourse connectives are some of the clearest indicators of discourse structure, annotating the arguments of the relations they convey provides information both about those arguments and about the range of possible discourse structures. In order to use information about discourse connectives to parse discourse structure, we need to know for any particular discourse connective whether it is structural or anaphoric. In order to make this designation, a systematic empirical study which shows the behaviour of the connective over a significant number of cases is required. In addition, for anaphoric connectives, in order to develop a resolution algorithm, symbolic or statistical, which can identify anaphoric arguments, a corpus study which provides evidence for patterns of location and properties of anaphoric arguments is a necessary first step. The corpus study undertaken here is very exploratory, but its general goal is to provide evidence to characterize particular connectives as structural or anaphoric, and if the latter, identify features characteristic of their anaphoric arguments.

3 *Corpus Study*

This work is a subset of a larger discourse annotation project whose main goal is to create a large corpus reliably annotated for discourse structure for further scientific research and development of NLP applications (e.g. question-answering, text summarization) (Miltsakaki, *et al.* 2004). Each overt or null discourse connective in the corpus will be marked with the minimal textual unit in the preceding discourse which contains the source of its left-hand argument. Although for the purposes of both the present study and the larger corpus study a strictly textual antecedent is being identified, this is a practical simplification of the theory. In fact, the anaphoric argument is more accurately treated as an Abstract Object (AO) in the discourse model, the same kind of entity that can be accessed through a demonstrative pronoun (or *discourse deictic*), as argued

represent pronominal reference (Heim & Kratzer, 1998), i.e. an pronoun denotes an entity e via an index i that is mapped to e relative to an assignment function a , where a is determined by a context c (e.g. $[[\text{you}_i]]^{ac}$, where c might yield *Tom, Dick, Harry* for i).

for in Forbes (2003).⁷ However, in light of the fact that many successful current approaches to anaphor resolution of NPs, apply to surface elements, from an engineering perspective, identification of the textual material which gives rise to the AO is a more realistic task. The success of the overall project will contribute to our ability to understand and deal with an important aspect of discourse meaning, i.e. discourse relations.

3.1 Corpus Annotation

The work we report here is a first attempt to taxonomize the set of discourse connectives and their properties. To the best of our knowledge, annotation of the arguments of discourse connectives has not been previously attempted. As such, the annotation here is necessarily experimental and explorative, and to some extent the annotation guidelines were developed and altered as the annotation took place. We began with a set of nine connectives picked from three semantic classes: resultatives (*as a result of, so, therefore*), concessives (*nevertheless, yet, whereas*), and additives (*also, in addition, moreover*). They are all adverbials, which may, by definition, modify phrasal constituents (ADJP, PP, VP) or an entire clause.^{8,9}

For each of the nine connectives, seventy-five tokens (for a total of 675 tokens) were extracted from a variety of corpora: Brown, Wall Street Journal, Switchboard and 58 transcribed oral histories from the online Social Security Administration Oral History Archives (SSA).¹⁰ The 675 tokens were split in three groups (each group containing a connective from each semantic class) and annotated by three annotators (225 tokens per annotator).

Each token was annotated with tags that encoded information about (a) the connective's left argument (ARG), and (b) the clause containing the connective (CONN). Table 1 shows the ARG and CONN tag(sets) in the top and bottom box respectively. Both ARG and CONN were annotated with a REF tag that encoded an ID number which linked the two parts of the single token. ARG was

⁷ Note also that, just as with bridging reference for NPs, the argument of a discourse adverbial may be an abstract object derived from, but not identical to, an AO already in the discourse model (Webber *et al.*, 2003).

⁸ In other words, lexical items that function as connectives also have other syntactic roles at the sentential level (e.g. *he is otherwise occupied, hereafter happy to eat tofu, so tired*, etc.). This study excludes these other uses on the grounds that they must be accounted for as part of sentential syntax.

⁹ Although *whereas* can be used as an adverbial connective, in our data it mostly appears as a subordinate conjunction because a clause introduced by it can appear before or after its other clausal argument, as shown in example (1).

(1) Whereas persons of eighth grade education or less were more apt to avoid or be shocked by nudity, those educated beyond the eighth grade increasingly welcomed and approved nudity in sexual relations. (Brown)

¹⁰ The Brown, Wall Street Journal and Switchboard corpora are available from LDC, <http://www.ldc.upenn.edu>. The SSA corpus is available at <http://www.ssa.gov/history/orallist.html>

further tagged with a TYPE tagset that identified the extent of the argument. The tags under TYPE were as follows: MAIN if the argument was contained in a full sentence (including subordinate clauses); MAIN-MULT if the argument was contained in a sequence of sentences; SUB if the argument was contained in a subordinate clause; and XP if the argument was contained in a phrasal constituent. The variation in the size of the argument was thus specified as a structural description.

In the TYPE tagset, two additional tags were added during the annotation. The category OTHER was added in order to describe cases where the left argument of the connective could not be identified. The category NONE was added for cases where both arguments are to the right of the connective, and therefore, there is no left-hand argument. This tag applies only to cases of subordinate conjunctions, and so it only appeared in the annotation of *whereas* here.

The set of tags used for the type of the left-hand argument were selected in order to enable us to identify statistically useful information about the type of the antecedent of anaphoric connectives, which will ultimately allow the selection of features for use in a statistical or a symbolic anaphora resolution algorithm. In particular, the distinction between MAIN/MAIN-MULT and SUB/XP combined with the LOC tag (discussed in Section 3.3) will help determine optimal structural descriptions for the connectives that will be useful for systems like the DLTAG parser (Forbes *et al.*, 2003). For example, connectives found to take only contiguous MAIN/MAIN-MULT arguments can be associated with a tree taking two structural arguments, thus maximizing compositional semantic representations derived directly from the syntax of discourse.

The clause containing the connective, CONN, was annotated with two tagsets: COMB and POSITION. The tag COMB was used to identify punctuation marks (PERIOD, COMMA, etc.), coordinating conjunctions ('AND' and 'BUT'), and adverbial connectives ('YET', 'SO', etc.) that can co-occur with the connective.¹¹ Information about co-occurrence with punctuation

¹¹ Because all the adverbial connectives were annotated separately, in cases where *so* occurred with one of the adverbial connectives in the study, the token could be annotated both as an instance of that adverbial connective and as an instance of *so*. However, because *so* is a much more frequent connective than any of the adverbials it can potentially combine with, the first 75 instances of adverbial connective *so* that were encountered and used for the study did not include any cases where it combined with another adverbial (i.e. there was no overlap in the set of tokens of *so* annotated as *so* and the set of tokens of other connectives where *so* appeared; the latter were annotated as instances of the other connective.) There were five cases where *therefore* appeared with *so*. Here the effects of *so* and *therefore* on the location of the left-hand argument can be somewhat distinguished by comparing the location of the argument in the actual token with the location of the argument in an example identical but for the absence of *so*. The exploratory

and other (mainly structural) connectives will also be useful for determining syntactic properties of connectives. In DLTAG, *and* and *but* are structural connectives anchoring elementary trees. That is, both their arguments must be realized structurally. Co-occurrence with *and* and *but* may be an indication that a connective cannot take both its arguments structurally because such a structure would be underivable¹² or would require the assignment of computationally complex structural descriptions.

The tag ZERO in the COMB tagset is primarily relevant for tagging tokens of *whereas*. It describes cases where the conjunction combines with no punctuation marks or other connectives. However, in most cases, the presence of this tag indicates that the connective is a subordinate conjunction. Subordinate conjunctions do not combine with a punctuation mark – because of punctuation conventions in written English – or other connectives when the subordinate clause appears after the main clause. The ZERO tag applies much less frequently to adverbial connectives, like *also*.

Finally, we found it useful to make special tags for combinations with a complementizer (COMP) and a subordinate conjunction (SUB) because several connectives appear in complement and subordinate clauses. This creates ambiguity in their interpretation, discussed below in Section 4.1.

For the purposes of anaphora resolution, co-occurrence with punctuation combined with the results of the argument-size (i.e. TYPE) annotation will be useful features in guiding an automated search for anaphoric arguments. Also, certain types of punctuation, e.g. dashes and parentheses, may indicate that the text containing the argument of the connective is not adjacent to the clause containing the connective.

Co-occurrence with other connectives also raises the question of the semantics of the combined connective and its relationship to the semantics of the individual contributors, as for example, in the combinations *and in addition* or *yet nevertheless*.

For CONN, we also defined a POSITION tagset which identified the position of the connective in its clause (INITIAL, MEDIAL, FINAL). As

nature of the present study did not allow for full investigation of these effects, but because they were part of the annotation schema, they can be given more attention in a future study.

¹² In other words, the combination of two structural connectives (i.e. appearance within a single clause) cannot be derived under the current framework. This could be an empirically desirable result if there is additional, separate evidence supporting the inability of combining two structural connectives. On the other hand, if independent evidence for characterizing two connectives as structural exists and these connectives can be combined, then possibly the formal framework may have to be altered. The results of this study appear to favor the latter conclusion because *so* appears to be a structural connective, and it can clearly co-occur with *and*. The necessary revision of the formal framework to account for this phenomenon and/or more detailed investigation of the behavior of *so* remain for future work.

mentioned above, the position of the connective in the clause is relevant for the information structure partitioning of the clause (Forbes *et al.*, 2003; Kruijff-Korbayová & Webber, 2001), and thus this is useful information to keep track of with respect to particular connectives for parsing purposes.

A sample full annotation of an instance of *therefore* is shown in example (6). Here the left-hand argument, or ARG, is a main clause (TYPE=MAIN) that immediately precedes the sentence which contains *therefore*. *Therefore* itself appears medially (POSITION=MEDIAL) in a clause introduced by *and* (COMB=AND).

- (6) <ARG REF=27 TYPE=MAIN> Philip Lee was the Chancellor of the campus at San Francisco </ARG>
 <CONN REF=27 COMB=AND POSITION=MED> and he was **therefore** the person who hired me for the post as Director of the Medical Center. </CONN >

The complete set of tags we defined is given in Table 1, and an example of each tag is provided in examples (7–9).

ARG:	REF	ID #
	TYPE	MAIN = sentence MAIN-MULT = multiple sentences SUB = subordinate clause XP = phrasal constituent OTHER = no argument NONE = no <i>left</i> argument
CONN:	REF	ID #
	COMB	PERIOD COMMA COLON SEMI-COLON DASH 'AND' 'BUT' 'YET' 'SO' ZERO COMP SUB
	POSITION	INITIAL MEDIAL FINAL

Table 1: Annotation tagsets

- (7) a. MAIN: <ARG>On the basis of the applications, social security cards had been issued to people</ARG> <CONN> and two records therefore came to Baltimore.</CONN> One was the application form, the SS-5; the other was the office record form. (SSA)
 b. MAIN-MULT: Well, John Corson as assistant executive director was a wonderful, wonderful foil for Frank Bane, because Frank Bane never likes to say “no” to anybody,

- you know, even the executive director. And he could always say, “Yes,” when the answer was yes. <ARG>But when the answer was, “No,” John Corson would always give the answer. And John, knowing that this was the role for the second man, would handle it,</ARG> <CONN> and therefore all the onus that built up in the organization when a bureau director or staff member didn’t get what he wanted fell on John and not on Frank.</CONN> (SSA)
- c. SUB: And then these people would argue <ARG> we no longer need that sort of effect</ARG> <CONN> and, therefore, we don’t need a retirement test any more. </CONN> (SSA)
- d. XP: Of course, the contractors were to be <ARG> out there, </ARG> <CONN> and therefore part of the field, </CONN> (SSA)
- e. OTHER: *Claims precedent lacking*. After reading his statement discharging the 23d ward case, Karns told Wexler that <CONN> if the seven cases scheduled for trial also involved persons who had been subpoenaed, </CONN>he would dismiss them. (Brown)¹³
- f. NONE: <CONN> Whereas most men were a bit ambivalent about the sex scandals (though they were furious about Recruit), </CONN> <ARG> women were upset about both and surged to the polls. </ARG> (Brown)
- (8) a. PERIOD: Well that gave me sort of an insight, so I made it a practice to contact all of the funeral directors, which in those days was forbidden. **Nevertheless**, I went ahead and contacted them anyway. (SSA)
- b. COMMA: Although Sam Rayburn affects a gruff exterior in many instances, **nevertheless** he is fundamentally a man of warm heart and gentle disposition. (Brown)
- c. SEMI-COLON: I am thoroughly convinced that most watercolors suffer because the artist expects nature will do his composing for him; **as a result**, such pictures are only a literal translation of what the artist finds in the scene before him. (Brown)
- d. DASH: The 1954 Amendments completely changed the financing of the vocational rehabilitation program, providing for a three-part grant structure – for (1) basic support; (2) extension and improvement; and (3) research, demonstrations, training and traineeships for vocational rehabilitation — and **in addition** for short-term training and instruction. (SSA)
- e. ‘AND’: But it is still a quasi-Independent Agency **and therefore** the ability to be able to speak one’s mind is certainly more than it is for traditional cabinet-level officials or senior political officials who serve at the pleasure of the President. (SSA)
- f. ‘BUT’: **But nevertheless** consultation is the prime instrumentality that you use to get support. (SSA)
- g. ‘SO’: **So therefore**, if you have some situations that arise when maybe an ALJ put someone on that the DDS didn’t think was disabled, you’ve got to show the person improved over what the ALJ said before you can take the person off. (SSA)
- h. ‘YET’: This, plus the habit of many schools of simply adding interior design to the many subjects of their home economics department, **yet, nevertheless**, claiming that they teach interior design, has contributed to the low repute of many university courses in interior design. (Brown)
- i. ZERO: The Controller’s charge of rigging was the latest development in an investigation which **also** brought these disclosures Tuesday : [...] (Brown)

¹³ The missing argument here is roughly *The 23rd ward case involved persons who had been subpoenaed*. This proposition is not expressed explicitly anywhere in the article.

- j. SUB: After the first few weeks, it was obvious that rules had to be made, laid down and obeyed — even if our popularity ratings became subnormal **as a result**. (SSA)
 - k. COMP: Moritz said Monday that his leg feels fine and, **as a result**, he hopes to start practicing field goals this week. (Brown)
- (9) a. INITIAL: **Nevertheless** he had ample opportunity to contest the statement before the appeal board. (Brown)
- b. MEDIAL: Only those who were actually investors, **therefore**, were eligible for a lump-sum return at reaching age 65 or the widow would receive it at his death. (SSA)
 - c. FINAL: But it is true, **nevertheless**. (Brown)

Although each tag was explained and illustrated with examples like those above, each annotator was guided wholly by their intuition when determining the values of each tag for each anaphoric argument they annotated. Below in Section 4, we discuss how this intuitive guideline can be further refined: by studying patterns that emerge across similarities and differences between all the annotators’ intuitive decisions, we develop a set of heuristics that both improve the guidelines and improve the inter-annotator agreement.

3.2 Annotation Results

Table 2 shows the results of the preliminary annotation for the nine connectives. The table contains percentages of the tags TYPE and POSITION along with the actual number of occurrences of the tags in brackets. In the COMB tagset, a connective could combine with more than one of the categories of the group, so no percentages are given because the numbers do not add up to 75 for each category.

For most connectives, there is a strong tendency for the left argument to be identified locally (in the structural sense) – either in the immediately preceding sentence or in an immediately preceding sequence of sentences, (e.g. the preceding paragraph).¹⁴ Most notably, *so* always takes a sentence or a sequence of sentences (i.e. a segment made up of multiple sentences) immediately preceding it as its left argument, indicating that it may tentatively be treated as a structural connective. *In addition*, *yet*, *moreover*, *as a result* and *also* tend to take their left argument locally but they demonstrate a larger syntactic variety of potential arguments such as subordinate clauses or phrasal constituents. *So*, *nevertheless* and *moreover* are more likely than the others to take larger discourse segments as arguments, adjacent in the case of *so* and not necessarily

¹⁴ No limited window was set as a search space for a potential argument. This allowed annotators to look as far back as needed in a particular text to find the location of the argument. In a few rare cases of the 675 tokens examined in the first part of this study, the left argument spanned multiple paragraphs. As can be seen from the results to be presented in Section 3.3, left arguments non-contiguous with their connective are also exceptional. From this, we can conclude that in future development of resolution algorithms, setting the window to be examined to at most the paragraph containing the connective and one or two preceding paragraphs would not harm accuracy greatly.

adjacent in the case of *nevertheless* and *moreover*. The connective *therefore* often takes its left-hand argument from a subordinate clause. Larger discourse segments appear to lead to vagueness in resolving anaphora (cf. Section 4). For example, it was often difficult to determine the extent of a multi-sentence left-hand argument of *nevertheless*. *Nevertheless* can also find its anaphoric argument in an intra-sentential constituent (XP).

Regarding the position of connectives, *so* appears only in initial position. This supports the claim that *so* is a structural connective because the quintessential structural connectives — subordinate and coordinate conjunctions — are restricted to the initial position. *Also*, on the other hand, frequently appears in medial positions, while the semantically similar *in addition* prefers the initial position.

Connective	in addition	so	yet	nevertheless	moreover	therefore	as a result	whereas	also
Type									
Main	65.3% (49)	45.0% (34)	53.3% (40)	37.3% (28)	42.7% (32)	25.3% (19)	78.6% (59)	46.7% (35)	69.3% (52)
Main-Mult	18.7% (14)	55.0% (41)	33.3% (25)	36.0% (27)	45.3% (34)	21.3% (16)	18.7% (14)	4.0% (3)	9.3% (7)
Sub	5.3% (4)	0.0% (0)	2.7% (2)	9.3% (7)	8.0% (6)	31.0% (24)	2.7% (2)	16.0% (12)	12.0% (9)
XP	10.7% (8)	0.0% (0)	10.7% (8)	17.3% (13)	4.0% (3)	21.3% (16)	0.0% (0)	1.3% (1)	4.0% (3)
(none)	--	--	--	--	--	--	--	32.0% (24)	--
(other)	--	--	--	--	--	--	--	--	5.3% (4)
Comb									
Period	65	33	33	47	68	28	55	26	49
Comma	9	22	14	5	2	1	0	36	7
Semicolon	1	2	8	0	0	0	3	5	0
Dash	1	0	4	0	0	0	0	0	0
And	12	2	8	1	4	41	14	0	7
But	0	0	0	17	1	0	1	0	4
Yet	0	0	0	2	0	0	0	0	0
So	0	0	0	0	0	5	0	0	0
Zero	0	0	0	3	2	0	0	8	1
Comp	0	0	0	0	0	0	1	0	8
Sub	0	0	0	0	0	0	1	0	0
Pos									
Initial	92.0% (69)	100.0% (75)	98.7% (74)	78.6% (59)	82.7% (62)	88.0% (66)	90.7% (68)	100.0% (75)	17.3% (13)
Medial	8.0% (6)	0.0% (0)	1.3% (1)	18.7% (14)	17.3% (13)	12.0% (9)	2.7% (2)	0.0% (0)	80.0% (60)
Final	0.0% (0)	0.0% (0)	0.0% (0)	2.7% (2)	0.0% (0)	0.0% (0)	6.6% (5)	0.0% (0)	2.7% (2)

Table 2: Annotation results for 9 connectives, 75 tokens each

The results of this initial annotation project are promising because they reveal interesting variation in distribution patterns. To further revise the annotation tags and guidelines and, crucially, test inter-annotator reliability, we focused our attention on three connectives *as a result*, *in addition* and *nevertheless*, one from each of the three semantic classes. Another twenty-five tokens of each of the three connectives were extracted to add up to a total of one hundred per connective and give an indication of intra-annotator consistency. The annotation of the complete set of three hundred tokens for the three connectives appears in Table 3. Comparison of Tables 2 and 3 shows that the relative percentages of each tag remained stable, indicating that the anaphoric arguments of each of these connectives display patterns that can be recognized via a large-scale annotation project, and be used to lead to reliable

annotation algorithms. What remains to be shown is that this annotation is reliable, such that the same patterns are perceived across annotators.

Connective	In addition	Nevertheless	As a result
Type			
Main	63% (63)	36% (36)	68% (68)
Main-Mult	19% (19)	35% (35)	26% (26)
Sub/Comp	10% (10)	10% (10)	5% (5)
XP	8% (8)	18% (18)	0% (0)
Other	0% (0)	0% (0)	1% (1)
Comb			
Punctuation	101	78	80
Dash	1	0	0
And	12	1	17
But	0	2	1
Conn	0	2	0
Comp	0	0	10
Sub	0	0	1
Pos			
Initial	94% (94)	82% (82)	91% (91)
Medial	6% (6)	16% (16)	3% (3)
Final	0% (0)	2% (2)	6% (6)

Table 3: Annotation results for 3 connectives, 100 tokens of each

3.3 Inter-Annotator Agreement

Our studies in the prior section suggest that a human can identify and find patterns in the arguments of the connectives studied. The study presented in this section suggests that this identification and the patterns found are reliable. To test the reliability of our annotation, three additional annotators annotated 25 of the original 100 tokens of each of the three connectives (*in addition*, *as a result*, *nevertheless*), yielding a total of four annotations of 25 tokens of each of these connectives. Each connective and its anaphoric argument were, as in the prior study, assigned an ID. However, in order to focus on the ability of multiple annotators to agree on the unit from which the anaphoric argument is derived, we employed only one tag, LOC. Each annotator labelled the anaphoric argument with one of the four possible values of this tag shown in Table 4.

LOC:	SS = same sentence PS = previous sentence PP = previous paragraph NC = non-contiguous
------	--

Table 4: Values for argument tag LOC

The LOC tag defines the sentence as the relevant atomic unit from which the anaphoric argument is derived. A *sentence* is minimally a main clause and all

(if any) of its attached subordinate clauses. The semantic argument of the connective could thus be derived from the single sentence containing the connective (SS), the single prior sentence (PS), a sequence of adjacent sentences (PP), or a sequence of sentences not contiguous to the clause containing the connective (NC). In other words, we did not ask the annotators to distinguish sub-clausal constituents or subordinate clauses; we did not distinguish the exact boundaries of sequences of sentences when we marked more than one sentence as the argument; and we did not distinguish whether a non-adjacent argument comprised one clause or a sequence of them. In this sense, the LOC tag can be viewed as a generalization of the TYPE tag; however, it adds the additional information of whether the anaphoric argument is contiguous to the clause containing the connective. Reasons for employing the LOC tag will be discussed in Section 4.

The Table in the Annex shows the annotations for each set of 25 connective tokens using the LOC tag. The first column indicates the token number being annotated. Then, for each inter-annotation, the first four columns contain the particular LOC tag given to that token by each annotator, and the fifth column shows the proportion of annotators who agreed on a LOC tag for that token, i.e. 4/4 represents the case in which all four annotators produced the same tag, 3/4 represents the case in which three out of four annotators produced the same tag, 2/4 represents the case where two out of four annotators produced the same tag but the remaining two annotators had different tags, and <2,2>/4 represents the case where two annotators produced one tag, and the other two annotators produced another tag.

A summary of the inter-annotator results for the 25 tokens for these three connectives produced using the LOC tag is shown in Table 5. The first column indicates the connective, and the remaining columns contain the percentage of tokens in which a particular pattern of agreement was found for each connective. Again, the first column (4/4) represents the case in which all four annotators produced the same tag, the second column (3/4) represents the case in which three out of four annotators produced the same tag, the third column (2/4) represents the case where two out of four annotators produced the same tag but the remaining two annotators had different tags, and the fourth column (<2, 2>/4) represents the case where two annotators produced one tag, and the other two annotators produced another tag. That there is no “0” column reflects the fact that in every case, there was some agreement among annotators, e.g. there was no case in which each annotator selected a different tag.

Connective	4/4	3/4	2/4	<2,2>/4
in addition	76% (19)	16% (4)	4% (1)	4% (1)
as a result	84% (21)	12% (3)	0	4% (1)

nevertheless	52% (13)	36% (9)	0	12% (3)
--------------	----------	---------	---	---------

Table 5: Inter-annotator agreement by raw percentages

As Table 5 shows, four-way inter-annotator agreement is greater than 50% in every case, and majority agreement (three-way or better) is 92% for *in addition*, 96% for *as a result*, and 88% for *nevertheless*. Inspection of the individual annotations in the Table in the Annex further demonstrate that the annotators almost always agreed on the use of the SS tag. In other words, the annotators were in agreement when distinguishing anaphoric arguments in the same sentence as the connective from anaphoric arguments farther back in the discourse. The most difficult distinction found across all the connectives concerned whether the anaphoric argument was contained in the prior sentence (PS) or some larger chunk of the prior contiguous discourse (PP). This table also shows that the anaphoric argument was almost always agreed to be contiguous to the clause containing the connective, i.e. the NC tag was rarely used.

Tables 6-8 break down these inter-annotation agreement results by pairs of annotators, using the Kappa statistic. Kappa values are used to measure the degree to which two annotators concur in their respective sortings of N items into k mutually exclusive categories. In the present study, 25 tokens are sorted into one of 4 categories, represented by the 4 values of the LOC tag.¹⁵ Note that these tables show Cohen's unweighted Kappa value for each pair of annotators, for each connective, e.g. the value located in the third row and fourth column of Table 6 shows that the annotations of ANN_k and ANN_c had a Kappa value of 0.88.

One can alternatively compute weighted Kappa values, and this may in fact be more appropriate to this study; however, weighted Kappas require that one can accurately determine how to weight each category. For unweighted Kappa, category weightings are by default set to '1'. Alternative weightings can be determined by the imputed relative distances between categories. At this point, we use unweighted Kappa because determining how to weight each of our LOC tags is an unresolved empirical question. It may be, for example, that there is a tendency to prefer the closest likely discourse unit over others farther away in the discourse as the anaphoric argument. We discuss such issues further in the next section, but the overall question is still an open one.

	ANN _c	ANN _k	ANN _c	ANN _r
ANN _c	--	0.61	0.66	0.74
ANN _k	--	--	0.88	0.81
ANN _c	--	--	--	0.81

¹⁵ See Carletta (1996) for details on computing Kappa values.

Table 6: Kappa values for *in addition* annotation across 4 annotators

	ANN _e	ANN _k	ANN _c	ANN _r
ANN _e	--	0.67	0.76	0.84
ANN _k	--	--	0.91	0.74
ANN _c	--	--	--	0.83

Table 7: Kappa values for *as a result* annotation across 4 annotators

	ANN _e	ANN _k	ANN _c	ANN _r
ANN _e	--	0.58	0.59	0.58
ANN _k	--	--	0.65	0.53
ANN _c	--	--	--	0.76

Table 8: Kappa values for *nevertheless* annotation across 4 annotators

As shown, Kappa values for the *in addition* annotation range from 0.61–0.88, and yield an average Kappa across the 4 annotators of 0.75. Kappa values for the *as a result* annotation range from 0.67–0.91, and yield an average Kappa across the 4 annotators of 0.79. Kappa values for the *nevertheless* annotation range from 0.53–0.76, and yield an average kappa across the 4 annotators of 0.62. Across all three connectives, Kappa values range from 0.53–0.91 and yield an average of 0.72.

Overall, both the raw percentages and the Kappa-statistic evaluations of our inter-annotation agreement reflect the fact that *nevertheless* was more difficult to annotate than either *in addition* or *as a result*. As the project expands, we will likely continue to find both more and less difficult annotation cases. Based on what we’ve seen so far, however, we conclude that the anaphoric arguments of discourse connectives can be reliably annotated.

In the next section, we discuss how investigating of annotator disagreements can be used to develop a resolution algorithm for the anaphoric arguments of discourse connectives.

4 Towards a Resolution Algorithm

A closer look at 1) how the annotations vary in the inter-annotator study and 2) the results of the more complex annotations in the individual annotation studies, reveals certain issues relevant to developing a resolution algorithm, including the need for a minimal argument heuristic, the existence of true ambiguity in identifying arguments, and the issue of whether anaphora resolution can guide decisions about parsing discourse structure.

4.1 Minimal argument heuristic

As mentioned above, we employed the LOC tag instead of the TYPE tag in the study of inter-annotator agreement. By additionally asking each annotator to record the boundaries of the units she identified as the “exact” unit from which

the anaphoric argument was derived, we were able to derive the values for the TYPE tags from each of the four annotations. For the purposes of inter-annotator agreement we found that “exact match” was not a useful comparison, due to differences in the implicit guidelines each annotator was individually following.¹⁶ However, the exact match comparison, combined with the data from the first study, is useful for elucidating these differences and understanding why they arise. The differences between the annotations fall into two main categories, the extent of the argument and the syntactic form of the argument. Both concern the annotator’s understanding of the properties of the unit that are necessary to derive the semantic argument of the connective.

Consider the discourse in (10). When deciding on the anaphoric argument of *as a result*, one annotator might decide that the decrease in blood pressure is the result of the decrease in stress and so tag the argument as PS. Because the decrease in stress about money is a result of Lee winning the lottery, however, another annotator might tag the argument as PP, e.g. as including both the first and second sentences.

- (10) Lee won the lottery. So, he was less stressed about money. As a result, his blood pressure went down.

Similarly, consider the discourse in (11). When deciding on the anaphoric argument of *as a result*, one annotator might decide that John’s being a man is the cause of his being drafted (females not being drafted in America historically), and thereby tag the argument as NC because John’s living in the US and registering for the selective service are an elaboration on the concept of being a male American. However, another annotator might tag the argument as PP, e.g. as including the first three clauses.

- (11) John is a male American. He has lived in the US his whole life. At 18, he registered for the selective service. As a result, he was drafted.

Finally, consider the sentence in (12). When deciding on the anaphoric argument of *as a result*, one annotator might decide that because *as a result* modifies an adjective on the right, its left argument should be (using the TYPE tag) an XP, e.g. *overworked*. Another annotator might interpret *tired* as a small clause, or a clause with a deleted subject and verb, and so he might tag the entire clause *Kim is overworked* as the anaphoric argument of *as a result* using

¹⁶ Because of the exploratory nature of the annotation project, the initial set of guidelines used for annotation were not detailed enough to prevent differences in annotation which would affect our ability to make use of string matching comparisons across annotators in any interesting way. For example, one annotator might systematically include punctuation or other connectives within an argument, while another excludes it.

the MAIN tag. (Note that this issue is avoided when the LOC tag SS is employed.)

(12) Kim is overworked, and as a result, tired.

What all of these cases have in common is the question of how large to make the argument. What they also have in common, however, is that in each case it is possible to select a minimal unit as the argument, and allow the relations between that unit and the surrounding context to complete the interpretation. In (10), if the annotator selects *So, he was less stressed about money* as the argument of *as a result*, the relation between Lee winning the lottery and being less stressed will not be lost because *so* will take as its anaphoric argument the semantic interpretation of *Lee won the lottery*.

Similarly, in (11) if the annotator selects only the clause *John is a male American* as the argument of *as a result*, the relation between John living in the US and registering for the selective service and John being drafted can still be recovered. The empty connective signalling basic elaboration will link the first two arguments to *John is a male American* structurally; their relation to John being drafted will be an indirect one through the resultative relation of John being drafted and John being a male American.¹⁷

An additional complication that arises in the annotation of examples like (12) is the role of the lower-level syntactic annotation. In the Penn Treebank, from which the majority of our data is drawn, there is no principled parsing of such cases, in that it is left to the annotator to decide whether a particular use of a gerund, adjective, etc., should be parsed as a clause with missing elements when it is modified by an adverbial discourse connective. Therefore, we cannot reliably invoke the syntactic parse to decide when to label the left argument as a clause or an XP. We could, however, draw an analogy with coordinating conjunctions, which are commonly parsed with two XP arguments (e.g. *Sue is happy and tired*), although at the semantic level, two propositions are arguably being conjoined. If we allow the syntactic XP unit to represent a full proposition in the semantics, then we can invoke the minimal unit heuristic here too. The annotator could be instructed to choose the smallest possible unit as the argument, in (12) the AdjP *overworked*, and then the full prepositional content of the argument, *Kim is overworked*, could be extracted from the sentential syntax and semantics. This would have the additional benefit of retaining parallelism in the surface syntactic form of the arguments of the connectives in such constructions.

¹⁷ Note that these same issues arise for a series of elaborations followed by *in addition*, and in the same way a minimal unit can be selected, under the assumption that the remainder of the connections can be reconstructed through all the links between minimal units.

Another potential heuristic in resolving the arguments of anaphoric connectives is their ability to combine with particular structural connectives, such as *but* and *and*. An auxiliary tree anchored with one of these connectives must be adjoined to its left-hand argument. Another connective, like *nevertheless*, *therefore*, or *moreover*, adjoined into this structure at the same point will frequently take the structural connective's left-hand argument as its own anaphoric argument (e.g. (13), where *and* and *therefore* share their left-hand argument).¹⁸

- (13) He believed that <ARG> the Federal Security administrator had the authority and the responsibility for actions taken throughout the agency, </ARG> <CONN>and therefore he should be apprised of them and should play a part in the decisions.</CONN> (SSA)

A similar heuristic could be used for determining the size of the left-hand argument. In particular, when the right argument is a constituent smaller than a full clause (e.g. the second of two conjoined VPs), the left argument appears to consistently be the same size (e.g. the first of two conjoined VPs), as in (14).

- (14) Jasper arrived late and therefore got no dinner.

An investigation of the variations in exact match labelling using the LOC tag and the individual labelling using the TYPE and COMB tags leads us to expect that if these heuristics are employed in the annotation, inter-annotator agreement will improve substantially. These minimal unit and connective combination cases, however, are distinguished from other issues that arise during the annotation of anaphoric arguments of discourse connectives, in that they are not cases of true ambiguity because principled heuristics can be introduced to resolve them. There are true cases of ambiguity, where such heuristics are not possible. One such case is discussed in the following section.

4.2 Ambiguity in Complement Clause

Cases of true ambiguity in identifying the left argument of a connective were found in connectives contained in complement clauses, mostly complements of verbs of saying. A connective in a complement clause may connect the complement clause with either the preceding sentence or with the main clause containing the verb of saying. To illustrate the point, consider example (15). This example is ambiguous between analyses (15a) and (15b).

- (15) Moritz said Monday that his leg feels fine and, as a result, he hopes to start practicing field goals this week.
 a. Moritz said Monday [that [his leg feels fine] [and, as a result, he hopes to start practicing field goals next week].]

¹⁸ But not always, as the examples, like that in (3) above, which motivate the distinction between anaphoric and structural connectives, demonstrate.

- b. [Moritz said Monday that his leg feels fine] [and, as a result, he hopes to start practicing field goals this week.]

In (15a), both arguments of *as a result* are embedded under *said*. The left argument is the first complement clause and is annotated as SS (same sentence) because both it and the connective clause are the conjoined object of the matrix clause verb. In (15b), the clause containing the connective is a main clause by itself. On this interpretation, *as a result* was not part of what Moritz said but was added by the writer. More generally, connectives appearing after a complement clause are ambiguous with respect to whether they are part of the embedded complement clause immediately preceding them (i.e. it is their left argument) or whether they are conjoined to the main clause (i.e. this higher clause is their left argument.)

4.3 Low Attachment

As stated above, one reason we used the LOC tag in inter-annotator agreement was because the TYPE tag did not distinguish contiguous from non-contiguous arguments. This is an important distinction to make, because such arguments cannot be modelled structurally, thus indicating that they must be resolved anaphorically.

Because anaphoric connectives do not retrieve their left argument structurally, the clause containing them must attach to the prior discourse tree via a tree anchored by an empty structural connective. The DLTAG parser (Forbes *et al.*, 2003) currently employs the procedure of always adjoining this empty connective tree to the leaf node on the right frontier of the growing tree. If the anaphoric argument could be identified through a resolution mechanism, the parser could use this information to decide to instead attach this empty connective to the node on the right frontier which dominates the anaphoric argument. This would mean that the resolution of the argument would in a sense be captured in the discourse structure tree.¹⁹ However, the anaphora resolution would have to precede the attachment decision here, so the structure cannot be thought of as in anyway determining the anaphora resolution. Moreover, examples like S₃ in the discourse in (3), show that this heuristic may very well not apply in cases where the anaphoric connective co-occurs with a lexical structural connective, rather than an empty connective.

¹⁹ The precise identity of the anaphoric argument would potentially remain ambiguous depending on the level where the anaphoric connective and its right argument were attached because that node might dominate several discourse segments.

5 *Conclusions and Future Work*

Discourse connectives are easily identified cues to discourse structure. But the actual discourse structure and relation that any particular connective indicates is not a fully-understood area of linguistic theory. By developing an annotated corpus of the discourse relations that individual connectives communicate through the anaphoric and structural connections they indicate, we can create an empirical picture of their behaviour which can be utilized in automatic detection of discourse structure.

We have reported the results of a preliminary corpus analysis of (primarily) anaphoric discourse connectives and the location and type of their arguments. The annotation provided information about whether particular connectives typically subcategorize for structural vs. anaphoric arguments. In addition, it provided detailed information about what the arguments look like and where they are found. This information will be useful for parsing discourse with a DLTAG. In addition, our results indicate that it will be possible to develop a resolution algorithm for identifying arguments that cannot be derived from the parse tree directly.

This study and the annotation guidelines developed as part of it are the starting point for a more extensive study which is creating a layer of annotations on top of both the Penn Treebank (syntactic) annotations and Prop Bank (semantic) annotations (Kingsbury & Palmer, 2002). Therefore, in the future we will be able to capture additional syntactic and also semantic properties of the sources of anaphoric arguments. These properties will be able to be automatically extracted from the annotated data. Additional annotation work on the discourse connective *instead* (Miltsakaki *et al.*, 2003) indicates that semantic properties of anaphoric arguments will be very useful for distinguishing them from non-arguments.

The annotation effort begun here, then, is a crucial first step towards automatic detection of the syntactic and semantic relations between propositions in discourse.

References

- Carletta, J. 1996. "Assessing Agreement on Classification Tasks: The Kappa Statistic." *Computational Linguistics* 22.
- Forbes, K. 2003. "Discourse Semantics of S-modifying Adverbials." Doctoral Dissertation, University of Pennsylvania.
- , E. Miltsakaki, R. Prasad, A. Sarkar, B. Webber and A. Joshi 2003. "D-LTAG System – Discourse Parsing with a Lexicalized Tree-adjoining Grammar." *Journal of Logic, Language and Information* 12, 261–279.
- Heim, I. and A. Kratzer 1998. *Semantics in Generative Grammar*. Oxford: Blackwell.

- Joshi, A., L. Levy and M. Takahashi 1975. "Tree Adjunct Grammars." *Journal of the Computer and System Sciences* 10, 136–163.
- Kingsbury, P. and M. Palmer 2002. "From Treebank to Propbank." *3rd Int. Conference on Language Resources and Evaluation (LREC-02)*. Las Palmas.
- Kruijff-Korbayová, I. and B. Webber 2001. "Information Structure and the Semantics of 'Otherwise'." In *ESSLLI'2001 Workshop on Information Structure, Discourse Structure and Discourse Semantics*, 61–78. Helsinki.
- Miltsakaki, E., C. Creswell, K. Forbes, A. Joshi and B. Webber 2003. "Anaphoric Arguments of Discourse Connectives: Semantic Properties of Antecedents versus Non-antecedents." *Proc. of the Computational Treatment of Anaphora Workshop, EACL 2003*. Budapest.
- , R. Prasad, A. Joshi and B. Webber. 2004. "The Penn Discourse Treebank." In *Proceedings of the Language Resources and Evaluation Conference*. Lisbon, Portugal.
- Webber, B., A. Knott and A. Joshi 2000. "Multiple Discourse Connectives in a Lexicalized Grammar for Discourse." R Muskens, H. Bunt and E. Thijsse (eds.) *Computing Meaning*, vol. 2 of *Studies in Linguistics and Philosophy Series*, 229–245. Dordrecht: Kluwer Academic Publishers.
- , M. Stone, A. Joshi, and A. Knott. 2003. "Anaphora and Discourse Structure." *Computational Linguistics* 29, 545—587.

Annex

Token	<i>In addition</i> inter-annotation					<i>As a result</i> inter-annotation					<i>Nevertheless</i> inter-annotation				
	Ann. _e	Ann. _k	Ann. _c	Ann. _r	Agreement	Ann. _e	Ann. _k	Ann. _c	Ann. _r	Agreement	Ann. _e	Ann. _k	Ann. _c	Ann. _r	Agreement
1	SS	NC	SS	SS	3/4	PS	PS	PS	PS	4/4	SS	SS	SS	SS	4/4
2	SS	SS	SS	SS	4/4	PS	SS	SS	SS	3/4	SS	SS	SS	SS	4/4
3	SS	SS	SS	SS	4/4	PS	PS	PS	PS	4/4	SS	SS	SS	SS	4/4
4	PS	PS	PS	PS	4/4	SS	SS	SS	SS	4/4	PS	PS	PS	PS	4/4
5	SS	SS	SS	SS	4/4	PP	PS	PS	PP	<2,2>/4	PS	PS	PS	PS	4/4
6	SS	SS	SS	SS	4/4	PP	PP	PP	PP	4/4	PS	NC	NC	PS	<2,2>/4
7	PS	PS	PS	PP	3/4	PP	PP	PP	PP	4/4	PS	PP	PS	PS	3/4
8	PS	PS	PS	PS	4/4	PP	PP	PP	PP	4/4	PP	PP	PP	PP	4/4
9	SS	SS	SS	SS	4/4	PS	PS	PS	PS	4/4	PS	PS	PS	PS	4/4
10	SS	SS	SS	SS	4/4	PS	PS	PS	PS	4/4	SS	SS	SS	SS	4/4
11	PS	NC	NC	NC	3/4	PS	PS	PS	PS	4/4	SS	SS	SS	SS	4/4
12	PS	PS	PS	PS	4/4	PS	PS	PS	PS	4/4	SS	SS	SS	SS	4/4
13	PS	PS	PS	PS	4/4	PS	PS	PS	PS	4/4	PP	PP	PP	PS	3/4
14	PS	PP	PP	PP	3/4	SS	SS	SS	SS	3/4	PP	PP	PS	PS	<2,2>/4
15	PS	PS	PS	PS	4/4	PS	PS	PS	PS	4/4	PP	PS	PP	PP	3/4
16	PS	PS	PS	PS	4/4	PS	PS	PS	PS	4/4	PP	PS	PP	PP	3/4
17	PS	PS	PS	PS	4/4	PS	PS	PS	PS	4/4	PS	PP	PP	PP	3/4
18	PS	PP	PP	PS	<2,2>/4	PS	PS	PS	PS	4/4	SS	SS	SS	SS	4/4
19	PS	PS	PS	PS	4/4	PS	PS	PS	PS	4/4	PS	PP	PP	PP	3/4
20	PS	PS	PS	PS	4/4	PP	PP	PP	PP	4/4	PS	PS	PS	PP	3/4
21	PS	PS	PS	PS	4/4	PS	PS	PS	PS	4/4	PP	PP	PP	PP	4/4
22	PS	PS	PS	PS	4/4	PP	PS	PP	PP	3/4	PS	PP	PP	PP	3/4
23	PS	PP	PP	PP	3/4	PS	PS	PS	PS	4/4	PS	PS	PP	PP	<2,2>/4
24	PP	PP	PP	PP	4/4	PP	PS	PS	PS	3/4	PS	PS	PP	PS	3/4
25	NC	PP	PS	NC	2/4	PS	PS	PS	PS	4/4	PS	PS	PS	PS	4/4