Reflections on the Penn Discourse TreeBank, Comparable Corpora, and Complementary Annotation

Rashmi Prasad[*]
University of Wisconsin–Milwaukee

Bonnie Webber[**]
University of Edinburgh

Aravind Joshi[†]
University of Pennsylvania

The Penn Discourse Treebank (PDTB) was released to the public in 2008. It remains the largest manually annotated corpus of discourse relations to date. Its focus on discourse relations that are either lexically grounded in explicit discourse connectives or associated with sentential adjacency has not only facilitated its use in language technology and psycholinguistics but also has spawned the annotation of comparable corpora in other languages and genres.

Given this situation, this paper has four aims: (1) to provide a comprehensive introduction to the PDTB for those who are unfamiliar with it; (2) to correct some wrong (or perhaps inadvertent) assumptions about the PDTB and its annotation that may have weakened previous results or the performance of decision procedures induced from the data; (3) to explain variations seen in the annotation of comparable resources in other languages and genres, which should allow developers of future comparable resources to recognize whether the variations are relevant to them; and (4) to enumerate and explain relationships between PDTB annotation and complementary annotation of other linguistic phenomena. The paper draws on work done by ourselves and others since the corpus was released.

1. Introduction

The Penn Discourse TreeBank, or PDTB (Prasad et al. 2008; PDTB-Group 2008) is the largest manually annotated resource of discourse relations. This annotation has been added to the million-word *Wall Street Journal* portion of the Penn Treebank (PTB) corpus

---

 * Department of Health Informatics and Administration, University of Wisconsin-Milwaukee, 2025 E. Newport Ave (NWQB), Milwaukee WI 53211. E-mail: `prasadr@uwm.edu`.
 ** School of Informatics, University of Edinburgh, 10 Crichton Street (IF4.29), Edinburgh UK EH8 9AB. E-mail: `bonnie.webber@ed.ac.uk`.
 † Institute for Research in Cognitive Science, University of Pennsylvania, 3401 Walnut Street (Suite 400A), Philadelphia PA 19104-6228. E-mail: `joshi@seas.upenn.edu`.

(Marcus, Santorini, and Marcinkiewicz 1993), indicating relations between the events, facts, states, and propositions conveyed in the text—relations that are essential to its understanding. Some relations are signalled explicitly, as in Example (1), where the underlined phrase <u>as a result</u> signals a causal relation between the situation described in the first two sentences (called *Arg1* in the PDTB, formatted here in *italics*) and a situation described in the third sentence (called **Arg2**, formatted here in **bold**). Other relations lack an explicit signal, as in Example (2), where there is no explicit signal of the causal relation between the situation described in the first sentence and that described in the second. Nevertheless, there is no change in meaning if the relation is made explicit—for example, using the same phrase <u>as a result</u> (Martin 1992).

(1) *Jewelry displays in department stores were often cluttered and uninspired. And the merchandise was, well, fake.*
     <u>As a result</u>, **marketers of faux gems steadily lost space in department stores to more fashionable rivals—cosmetics makers**. [wsj_0280]

(2) *In July, the Environmental Protection Agency imposed a gradual ban on virtually all uses of asbestos.* (implicit=<u>as a result</u>) **By 1997, almost all remaining uses of cancer-causing asbestos will be outlawed**. [wsj_0003]

Over 18K explicitly signalled relations and over 16K implicit forms have been annotated in the PDTB 2.0 (cf. Section 3.2, Table 1), which was released in February 2008, through the Linguistic Data Consortium (LDC).[1] Researchers since then, in both language technology and psycholinguistics, have begun to use the PDTB in their research, developing methods and tools for automatically annotating discourse relations (Wellner and Pustejovsky 2007; Elwell and Baldridge 2008; Pitler et al. 2008; Pitler and Nenkova 2009; Wellner 2009; Prasad et al. 2010a, 2011; Zhou et al. 2010; Ghosh et al. 2011a, 2011b, 2012; Lin, Ng, and Kan 2012; Ramesh et al. 2012), generating questions (Prasad and Joshi 2008; Agarwal, Shah, and Mannem 2011), ensuring an appropriate realization of discourse relations in the output of statistical machine translation (Meyer 2011; Meyer and Popescu-Belis 2012; Meyer and Webber 2013), and testing hypotheses about human discourse processing (Asr and Demberg 2012a, 2012b, 2013; Jiang 2013; Patterson and Kehler 2013). Other researchers have adapted the PDTB style of annotation to create comparable resources in other languages and genres (Section 4).

What then are the aims of this paper? First, for those researchers who are unaware of the PDTB, Section 2 of the paper lays out the key ideas behind the PDTB annotation methodology, and Section 3 describes the corpus in more detail than previous papers (Prasad et al. 2008; PDTB-Group 2008) and presents what we have learned since release of the corpus in 2008. Secondly, for those researchers who have used the PDTB, Section 3 aims to point out significant features of its annotation that have either been ignored or taken to be *intrinsic* when they are simply *accidental*. We hope that this will enable researchers to derive more from the corpus in the future and recognize the value of having it more completely annotated. Thirdly, annotation of comparable resources in other languages and genres has turned out to vary from PDTB annotation in ways that may be of interest to people contemplating the development of comparable resources in other languages and genres. Section 4 summarizes and explains the sources of this variation. Fourthly, Section 5 aims to show how PDTB annotation complements Time-Bank (Pustejovsky et al. 2003a) and PropBank (Palmer, Gildea, and Kingsbury 2005)

---

1 LDC Catalog ID LDC2008T05. `http://www.seas.upenn.edu/~pdtb` provides more information on the PDTB, including a complete list of publications.

annotation over the same Penn TreeBank corpus. Section 6 closes with a summary of the key points.

Although extensive documentation can be found on the PDTB Web site, along with discussions of various aspects of PDTB annotation, there has not as yet been as comprehensive and quantified a discussion of issues as presented here, especially concerning comparable corpora and complementary annotation. Providing this discussion is a major goal of this paper. We ourselves will be addressing many of these issues in the next few years.

2. Key Ideas Underlying PDTB Annotation

Two key ideas underlie the methodology used in annotating the PDTB, setting it apart from other efforts to annotate discourse relations (e.g., Carlson, Marcu, and Okurowski 2001; Polanyi et al. 2004; Baldridge, Asher, and Hunter 2007). First, it makes no commitment to any kind of higher-level discourse structure over the discourse relations annotated between individual text spans. Thus, while *theory-neutral* itself with respect to higher-level discourse structure, the PDTB invites experimentation with approaches to high-level topic and functional structuring (Stede 2012; Webber, Egg, and Kordoni 2012) or to hierarchical structuring (Mann and Thompson 1988; Asher and Lascarides 2003), as a resource for research aimed at a "data-driven and emergent theory of discourse structure" (Bunt, Prasad, & Joshi 2012, page 61).

Secondly, the annotation of discourse relations is *lexically grounded*. Rather than asking annotators to directly classify the sense of relations, which is a difficult task (Stede 2008), annotators were asked to look at lexical items that can signal discourse relations, such as the expression *As a result* in Example (1). When they did signal discourse relations, their arguments and senses were then annotated. Annotators were also asked to look at adjacent sentences that lacked one of these explicit signals. Where they inferred a discourse relation, they first labeled it with a lexical item that could serve as its explicit signal (such as *As a result* in Example (2)), before going on to classify its sense. In both cases, this lexical grounding was aimed at making the annotation more reliable, but it can also serve as a feature in the automated identification of discourse relations mentioned in Section 1.

A more detailed introduction to the PDTB can be found in the PDTB-2.0 overview paper (Prasad et al. 2008) and the PDTB-2.0 annotation manual (PDTB-Group 2008). Other papers describe specific aspects of the annotation such as the senses used in annotating relations (Miltsakaki et al. 2008), alternative lexicalizations (Prasad, Joshi, and Webber 2010b), and attribution (Prasad et al. 2007).

3. Key Features of PDTB Annotation

Here we discuss four key aspects of PDTB annotation that have been partially ignored or misunderstood: Explicitly signalled discourse relations; implicit discourse relations; properties of the arguments to discourse relations; and several issues concerning the senses of discourse relations and their annotation. These discussions extend the description of these features in the original PDTB overview paper and annotation manual.

3.1 Explicitly Signaled Discourse Relations

As Patterson and Kehler (2013) note, the inference of discourse relations may draw heavily upon world knowledge, but may also be facilitated by specific linguistic signals.

It is these signals that we discuss here, distinguishing between (1) the linguistic expressions that can explicitly signal a discourse relation; (2) the resource-limited subset of these expressions that were annotated as such in the PDTB; and (3) the consequences of this resource limit on annotation for using the PDTB.

We have taken the view that discourse relations hold between two and only two (possibly discontinuous) spans of text that can be interpreted as propositions, eventualities, beliefs, etc. (what Asher [1993] has called **abstract objects**). As such, the spans are primarily one or more sentences or clauses, and the expressions that can signal relations between them come from four well-defined syntactic classes:

- *Subordinating conjunctions:* because, although, when, if, as, etc.
- *Coordinating conjunctions:* and, but, so, nor, or (and paired versions of the latter — neither...nor, either...or)
- *Prepositional phrases:* as a result, in comparison, on the one hand...on the other hand, etc.
- *Adverbs:* then, however, instead, yet, likewise, subsequently, etc.

These we have called **discourse connectives**, or **explicit connectives**.

During the pilot phase of PDTB annotation, we took as explicit signals of discourse relations, linguistic expressions suggested by previous researchers (Halliday and Hasan 1976; Martin 1992; Knott 1996; Forbes-Riley, Webber, and Joshi 2006). This set was then enlarged as new connectives were found in the WSJ corpus itself. Also identified during this phase were productive modifiers of explicit connectives such as *apparently*, *at least partly*, *in large part*, *even*, *only*, and so on, which were then annotated as connective modifiers.[2]

What were not taken to be discourse connectives were adverbial *cue phrases*, including sentence-initial *Now* (Example (3)), *Well* (Example (4)), *So* (Example (5)), and *OK* (Example (6)), because they signal topic changes such as the beginning of a subtopic or a return to a previous topic (Hirschberg and Litman 1993), rather than relating particular discourse elements.

(3)   **Now** why, you have to ask yourself, would intelligent beings haul a bunch of rocks around the universe? [wsj_0550]

(4)   **Well**, mankind can rest easier for now. [wsj_1272]

(5)   **So**, OK kids, everybody on stage for "Carry On Trading." [wsj_2402]

(6)   When Mr. Jacobson walked into the office at 7:30 a.m. EDT, he announced: "**OK**, buckle up." [wsj_1171]

We did not intend to annotate as discourse connectives *pragmatic markers* such as *actually* and *in fact*, which serve to signal the conversational role of the speaker's matrix utterance—specifically, that it is "either aligned with or contrary to something previously said by another speaker, by the speaker on a previous occasion or to what people in general say" (Aijmer and Simon-Vandenbergen 2004). But *in fact* was annotated in the PDTB as a discourse connective, whereas *actually* was not. Nevertheless,

---

2 We have not yet seen much use made of these modifiers, even though it is clear that they can, for example, be used to disambiguate connectives. (E.g., **all** tokens of modified *ever since* convey a TEMPORAL sense, while only 51% of unmodified *since* do. Similarly, 76% of modified *even though* convey the sense CONCESSION, whereas only 37.3% of unmodified *though* do.)

this accidental annotation provides interesting information on what discourse relations *pragmatic markers* are associated with, which seems worth further study.

Resources then limited which types of explicit linguistic signals of discourse relations were actually annotated as such. In particular, sentence-initial prepositional phrases with an overt deictic argument (e.g., *for that reason*, *by then*) were not included in the set of explicit discourse connectives and hence not systematically annotated because it was felt this could be put off until deictic coreference was annotated more generally. The consequences of limiting *a priori* what were taken as possible signals for a discourse relation[3] meant that adjacent sentences lacking one of these expressions might contain a different sort of evidence for a discourse relation between them. The consequence for annotating implicit discourse relations is described in the next section.

3.2 Implicit Relations

The PDTB calls discourse relations that lack an explicit discourse connective between their arguments **implicit discourse relations**. Users of the PDTB thus need to understand (1) where and how implicit relations were annotated and (2) what was done in their absence.

As to the first point, the PDTB did not mandate unconstrained search for implicit discourse relations. Rather, annotators were asked to consider implicit discourse relations only between adjacent sentences within a paragraph, in the absence of an explicit connective relating them. The procedure involved (1) identifying one or more connectives that could be inserted between the two sentences without changing the discourse relation(s) between them, and then (2) specifying the sense of those relations. This had several consequences, each of which is discussed further in this section:

1. A sentence might bear no relation to its left-adjacent neighbor, even though a wider search might find some earlier text to which it was related.

2. Paragraph-initial sentences were taken to have no left-adjacent neighbor and were thus not examined as an argument to a discourse relation unless they contained an explicit discourse connective.

3. Implicit discourse relations were not annotated within a sentence except between clauses connected by a semicolon.

4. There were cases where annotators could not insert a connective between sentences because to do so appeared redundant.

5. There were cases where annotators could not insert a connective between sentences because they did not infer a discourse relation between them. Rather, the later sentence simply provided more information about an *entity* mentioned in the previous one.

6. Despite there being common patterns of multiple explicit connectives, annotators were not asked whether an implicit discourse relation might hold concurrently with a relation signalled with an explicit connective.

---

3 "Possible" because many expressions on the list have non-discourse functions as well—e.g., in addition to functioning as a discourse connective expressing a "result" relation, *so* can also function as an intensifier (*so short*) or as part of the verb phrase anaphor *do so*. Part of the annotation process involved excluding tokens that did not function as discourse connectives.

Table 1
Total relations annotated in the PDTB.

PDTB Relations	No. of tokens
Explicit	18,459
Implicit	16,224
AltLex	624
EntRel	5,210
NoRel	254
Total	40,600

Point (1) has been addressed in the BioDRB (Prasad et al. 2011), which adheres to most of the PDTB annotation conventions but allows an implicit discourse relation to hold between non-adjacent sentences within the same paragraph (cf. Section 4.5). This has reduced the proportion of potential implicit relations that were marked *NoRel* from 1.15% in the PDTB (254/22141, cf. Table 1) to 0.9% (29/3223) in the BioDRB (Prasad et al. 2011). The same choice was made in the Hindi DRB (Kolachina et al. 2012).

Points (2) and (3) remain gaps in PDTB annotation that we plan to address in the future. Example (7) illustrates point (3): One could insert a connective such as *afterwards* or *thereafter* before the free adjunct (i.e., *afterwards returning …*), making explicit the relation of temporal PRECEDENCE between the event expressed in the main clause and that in the free adjunct. (Arg1 and Arg2 are not indicated in italics and bold in Example (7) because free adjuncts have not yet been annotated in the PDTB.)

(7)   He flew to Fort Bragg, N.C., in September of that year for a course in psychological operations, returning to the School of the Americas in Panama for a two-month course called "military intelligence for officers." [wsj_2013]

In Section 5.2, we discuss how PropBank ArgM annotation can be used in addressing gaps in the annotation of sentence-internal implicit relations in the PDTB.

With respect to point (4), cases where inserting connectives seemed *redundant* were taken to arise from the relation being signalled by an expression from *outside* the set of explicit connectives. These expressions were annotated as *Alternative Lexicalizations* of evidence for discourse relations and their *Arg1* and **Arg2** annotated accordingly. We have counted them under *AltLex* relations in Table 1 rather than as *Implicit* discourse relations (also shown there). For example, in Example (8), inserting a connective like *because* between the sentences was felt to be redundant. Here, *One reason is* was annotated as an alternative lexicalization of the causal relation between them (indicated in SMALL CAPITALS adjacent to **Arg2**).

(8)   *Now, GM appears to be stepping up the pace of its factory consolidation to get in shape for the 1990s*. (CONTINGENCY.CAUSE.REASON) **One reason is mounting competition from new Japanese car plants in the U.S. that are pouring out more than one million vehicles a year at costs lower than GM can match**. [wsj_2338]

Some *AltLex* expressions are the deictic PPs which were not annotated as explicit discourse connectives due to resource limitations (cf. Section 3.1). Other expressions

926

such as *quite the contrary*, *eventually*, and *thereafter* (nearly 15% of alternative lexicalizations) meet all the criteria for explicit connectives, even though they had not been included earlier. Another 9% of expressions were found to be phrases such as *What's more* (Example (9)), which suggests that the range of discourse connectives should be widened to include other syntactic classes.

(9)   Marketers themselves are partly to blame: *They've increased spending for coupons and other short-term promotions at the expense of image-building advertising*. (EXPANSION.CONJUNCTION) **What's more, a flood of new products has given consumers a dizzying choice of brands, many of which are virtually carbon copies of one other**. [wsj_1856]

AltLex expressions are under-annotated in the PDTB because they were only annotated when annotators found it redundant to insert an implicit connective between adjacent sentences. For example, whereas 15 tokens of *that means* were noticed and annotated as *AltLex*, another 18 in the corpus were not examined, such as the one following *and* in Example (10).

(10)   "*I see a lot of evidence indicating a slower economy*, <u>and</u> **that means my interest-rate outlook has a downward tilt**," said Garnett L. Keith Jr. . . . [wsj_1694]

As a result, *AltLex* expressions cannot be exploited in machine learning—for example, for inducing a model of discourse relation annotation—because no individual *AltLex* expression can be guaranteed to be fully annotated in the corpus.

Everyone who has attempted to annotate or catalogue discourse connectives has commented on the lack of a complete list of words and phrases serving this role (Versley 2010; Rysová 2012; Meyer and Webber 2013). Rather than provide annotators with an incomplete list of connectives and allowing them to identify *alternative lexicalizations* during annotation), one might consider giving them complete freedom as to what to annotate as grounding for discourse relations. Although such a process has its own problems (Section 4.1), expert annotators and good training may make it a plausible basis for effective discourse annotation.

As for point (5), if annotators were unable to insert a connective between sentences because they were not able to infer a discourse relation between them, they were asked to check whether the second sentence provided more information about one or more entities mentioned in the previous sentence, as in

(11)   Pierre Vinken, 61 years old, will join the board as a nonexecutive director Nov. 29. Mr. Vinken is chairman of Elsevier N.V., the Dutch publishing group. [wsj_0001]

If it did, annotators were told to annotate the relation type as *EntRel*. EntRel captures **entity-based coherence** between sentences (Knott et al. 2001) realized either directly (i.e., via an anaphoric pronoun or NP in the second sentence) or indirectly (i.e., through a **bridging inference**). Annotators were not asked to annotate their evidence for EntRel. Although it might be possible to use a resource such as the coreference annotation in OntoNotes (Weischedel et al. 2012) to recover what the annotators had in mind, in cases such as Example (12) the *entity-based coherence* is less obvious.

(12)   *This financing system was created in the new law in order to keep the bailout spending from swelling the budget deficit*. **Another $20 billion would be raised through Treasury bonds, which pay lower interest rates**. [wsj_2200]

Such cases would benefit from the entity or entities that ground this relation being annotated explicitly, as well as provide valuable data for studies of entity-based coherence.

Table 1 also notes the 254 cases where the annotators did not see either an Alternative Lexicalization or an Entity-relation between adjacent sentences (within the same paragraph). These they annotated as NoRel.

Point (6) involves the possibility of an explicit connective or *AltLex* expression occurring concurrently with an implicit connective. That is, a recent unpublished pilot study carried out at the University of Edinburgh (Jiang 2013) used MTurk to show that readers presented with text containing a discourse adverbial also infer the sense associated with a conjunction (coordinating or subordinating), even when no conjunction is present in the text. The study involved 80 items taken from the freely available *Corpus of American English*, each consisting of a clause, followed by a gap, followed by a clause containing one of four discourse adverbials (*after all*, *in fact*, *in general*, *instead*). Each HIT (Human Intelligence Task) consisted of one item and six conjunctions (*and*, *because*, *before*, *but*, *or*, *so*, or none). The Turker was asked to insert into the gap, the conjunction (or none) that seemed most natural between the clauses.

For example, in Example (13), 50/52 Turkers inserted *but* into the gap, showing that they interpreted the relation between the clauses as being the same as if what was explicit was <u>but instead</u>:

(13)   Logically, she should be dead. <u>Instead</u>, she feels fine, caring for her
            daughters and walking a pedometer-measured two miles a day.

In Example (14), 49/52 Turkers inserted *so* into the gap, showing that they interpreted the relation as being the same as if what was explicit was <u>so instead</u>:

(14)   He suspected he shouldn't say that. <u>Instead</u> he lied.

Notice how different Example (13) would be if it were realized with <u>so instead</u>, or Example (14) with <u>but instead</u>: Neither is what the writer intended.

The effect was less strong in Example (15), where 33/52 Turkers chose <u>because</u>, showing that they interpreted the relation as being the same as <u>because instead</u>, whereas the other Turkers chose differently among the other options:

(15)   If he'd expected her to be upset, he was disappointed. <u>Instead</u>, she laughed,
            clapping her hands.

Although it is not yet clear which discourse adverbials are compatible with one or more concurrent implicit relations, it is nevertheless likely that such discourse relations are under-annotated in the PDTB and should be addressed.

3.3 Arguments

The two arguments to discourse relations contribute to the senses of the relations taken to hold between them. The PDTB gives annotators two ways to restrict these arguments to only the material needed for these relations.

As already noted, it is the events, states, propositions, claims, etc., in a text that participate in discourse relations. In English, such *abstract entities* tend to be conveyed through sentences, clauses, nominalizations, and verb phrases. Hence, these are what can be annotated in the PDTB as arguments.[4] Also because **discourse deictics** (e.g., *this*,

---

4 Because neither *and* nor *or* was annotated as a discourse connective when it conjoined VPs, so-conjoined
  VPs were not annotated as arguments.

that, *so*) can refer back to such interpretations (Example (16)), as can particles such as *yes* and *no* that function as responses to questions (Example (17)), these can also be annotated as arguments.

(16)   Evaluations suggest that good ones are—*especially so* if **the effects on participants are counted**. [wsj_2412]

(17)   Underclass youth are a special concern. [$_{Sup1}$ Are such expenditures worthwhile, then]? *Yes*, if **targeted**. [wsj_2412]

One way to limit arguments to only the minimal text needed for a given discourse relation (a *minimality principle*), was to allow annotators to specify that other text appeared *relevant but not necessary* to that interpretation. Specifically, they could annotate as *Sup1*, material *supplementary* to Arg1, as in Example (17), where the preceding question was annotated as relevant to interpreting the question response particle, and as *Sup2*, material supplementary to Arg2, as in Example (18), where the material enclosed in square brackets was annotated as relevant but not necessary to interpreting the temporal relation expressed with *then*.

(18)   *It acquired Thomas Edison's microphone patent* **and** then **immediately sued the Bell Co.** [$_{Sup2}$ claiming that the microphone invented by my grandfather, Emile Berliner, which had been sold to Bell for a princely $50,000, infringed upon Western Union's Edison patent.] [wsj_0091]

Supplementary information (both *Sup1* and *Sup2*) appears to have been under-annotated in the PDTB, mainly because annotaters were only invited, and not required, to check whether any text should be so annotated. This shows when one compares the number of *Sup1* or *Sup2* annotations on explicit discourse relations, which were annotated first, with the number of such annotations on implicit discourse relations, which were annotated on a subsequent pass: 1,571 explicit relations were annotated with supplementary information, whereas only 126 implicit relations were, despite nearly equal numbers of both. Before considering the existence of *Sup1* or *Sup2* as a feature indicative of the likely use of an explicit connective to signal a discourse relation (Patterson and Kehler 2013), it must be assessed whether this is an accidental feature of the PDTB's annotation or an intrinsic feature of the discourse relations themselves.

A second way of limiting arguments to only the minimal text needed to complete a given discourse relation involves the separate annotation of *attribution* (Prasad et al. 2007). This allows the attribution holding between an agent and an abstract object to be included or excluded from the discourse relation as appropriate. For example, in Example (19), annotators could exclude the attribution phrase "said Howard Rubel, an analyst with C.J. Lawrence, Morgan Grenfell Inc. in New York" from Arg1, as unnecessary for the discourse relation, while including the attribution phrase "Mr. Asman is also annoyed" as necessary for the discourse relation in Example (20).[5]

(19)   *Defense contractors "cannot continue to get contracts on that basis,"* said Howard Rubel, an analyst with C.J. Lawrence, Morgan Grenfell Inc. in New York. (implicit=because) **"The pain is too great."** [wsj_0673]

---

5 The PDTB also annotates attribution relations, capturing their textual signal and semantic features over each discourse relation and each of its arguments. For a full description of attribution and its annotation, the reader is referred to Prasad et al. (2007). Attribution is now being annotated as a separate layer over the WSJ (Pareti 2012), building on the PDTB attribution scheme, but aiming to capture the phenomena more comprehensively than in the PDTB.

(20)   *Mr. Asman is also annoyed that Mr. Castro has resisted collaboration with U.S.*
         officials, <u>even though</u> **by his own account that collaboration has been**
         devised essentially as a mechanism for acts directly hostile to the
         Cuban regime, such as facilitating defections. [wsj_1416]

Attribution differs from supplementary information in that, when its polarity is
negative, it can interact with discourse relations. (*Sup* has no such interaction.) This
can be seen by contrasting Example (21), where (negative) *denying* is part of **Arg2**, and
Example (22), where (negative) *denying* is not part of *Arg1*, but is rather its attribution.

(21)   *The U.S. wants the removal of what it perceives as barriers to investment*;
         (COMPARISON.CONTRAST) **Japan denies there are real barriers**. [wsj_0082]

(22)   Viacom denies *it's using pressure tactics*. (EXPANSION.RESTATEMENT.
         SPECIFICATION) "**We're willing to negotiate**," says Dennis Gillespie,
         executive vice president of marketing. [wsj_0060]

In Example (21), the *wanting* in *Arg1* is taken to contrast with the *denying* in **Arg2**. But
in Example (22), the negative polarity of *denying* as the attribution of *Arg1* means that
being willing to negotiate is taken to further specify **not** *using pressure tactics*.

These techniques are concerned with excluding material unnecessary to concluding
the existence of a particular discourse relation. There is no comparable attempt to
ensure that spans annotated as arguments to discourse relations include all the features
that motivate a given relation (Section 3.4). This can be seen with discourse relations
associated with the connective *instead*. Its *Arg1* must convey an alternative that does not
hold (Webber 2013): In Example (23), *Arg1* conveys that "a price for the new shares has
been set" is an alternative that does not hold.

(23)   *No price for the new shares has been set*. <u>Instead</u>, **the companies will leave**
         it up to the marketplace to decide. [wsj_0018]

But the features that allow an argument to convey an alternative that does not hold
may not be present in the argument itself. For example, the PDTB annotators agreed
that the clause "to be any silver lining" was *Arg1* of *instead* in Example (24), based on
the minimality principle mentioned at the start of this section. But there is nothing in
this argument that conveys that this alternative does not hold. That would require *Arg1*
to be "there isn't likely to be any silver lining." However, the annotators did not take
such an argument to be minimal.

(24)   In China, however, there isn't likely *to be any silver lining* because the
         economy remains guided primarily by the state.
         <u>Instead</u>, **China is likely to shell out ever-greater subsidies to its**
         coddled state-run enterprises, which ate up $18 billion in bailouts
         last year. [wsj_1646]

Although in the majority of cases, minimal argument spans do contain all the features
needed to license the annotated sense, this was not required by the PDTB guidelines.
This point has been missed in efforts to use the PDTB in training automated sense
recognition.

3.4 Senses and their Annotation

A well-known feature of the PDTB is its three-level hierarchy of senses (Figure 1). The
express purpose of this hierarchy was to allow back-off to a more general sense if (1) an
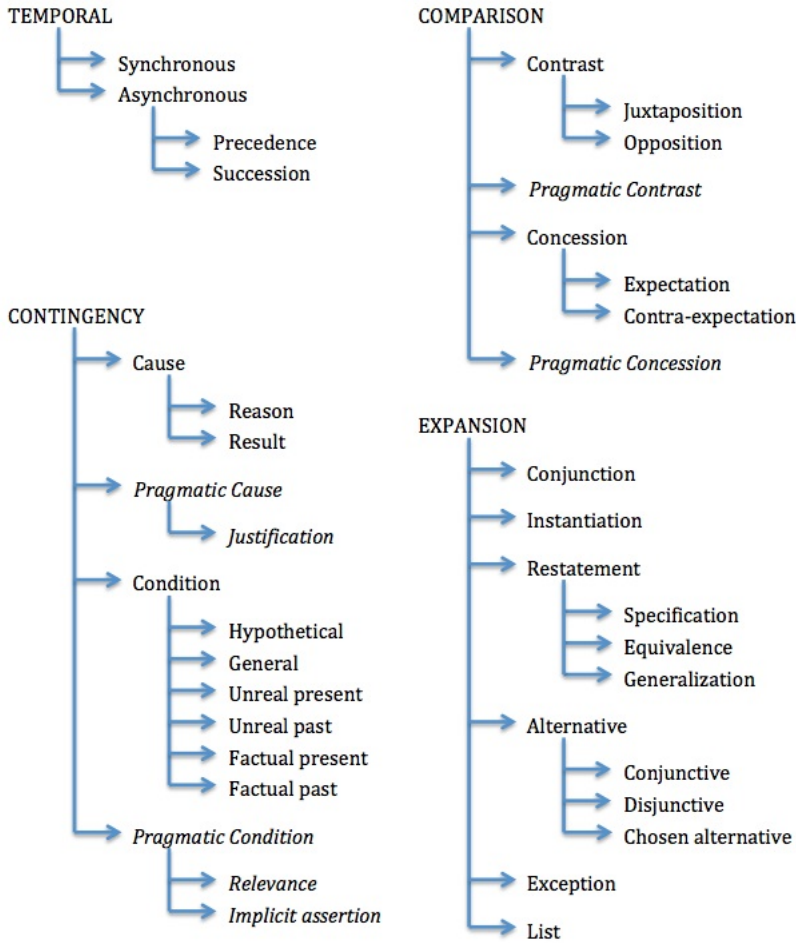
Figure 1
PDTB sense hierarchy.

individual annotator could not decide among its more specific senses or if (2) pairs of annotators disagreed as to a more specific sense.

Nevertheless, many researchers interested in inducing automated classifiers for explicit and/or implicit discourse relations have used the four top-level (level-1) sense classes for their research (e.g., Pitler et al. 2008; Pitler and Nenkova 2009; Zhou et al. 2010) because of the relatively large number of tokens in each class at this level of specificity (Table 2).

Our goal here is not to defend the hierarchy or its sense inventory, but rather to discuss three issues in sense labeling, which should help readers to better understand both the PDTB and the comparable corpora described in Section 4. The issues are: (1) senses found to be missing from the sense inventory; (2) disagreements between annotators; and (3) annotation of multiple concurrent discourse relations.

3.4.1 Missing Senses. We have long realized that there are gaps in the set of available senses. Some of these gaps are noted in the PDTB Annotation Manual (PDTB-Group 2008), such as the absense of a SIMILARITY sense for labeling explicit relations headed by *as if* and the absense of a PURPOSE sense for labeling explicit relations headed by

Table 2
Total explicit and implicit relations that fall under each level-1 sense.

level-1 PDTB Senses	No. of explicits	No. of implicits
CONTINGENCY	3,741	4,255
COMPARISON	5,589	2,503
EXPANSION	6,423	8,861
TEMPORAL	3,696	950

so that. Cases of the latter, as well as relations conveying the sense that one argument was the GOAL of the other, were simply annotated with the overloaded label CONTINGENCY.CAUSE.RESULT. The 34 cases of the subordinating conjunction *just as*, which can also signal SIMILARITY, were found on subsequent analysis of the corpus to be annotated as either TEMPORAL.SYNCHRONY (13 tokens) or EXPANSION.CONJUNCTION (1 token), or to have been left unannotated (20 tokens).

Some of these and other missing senses have been added to the sense inventories used in annotating corpora comparable to the PDTB (Section 4.2). They should also be added to the PDTB, and senses known to be overloaded should be split appropriately. Although this will eliminate already noted sense gaps, the potential remains for additional senses to be identified, and hence this remains a problem.

3.4.2 Disagreements Between Annotators. Sense annotation was done by two annotators. Disagreement at level-1 occurred when the two annotators picked senses that belonged at or under different level-1 classes. Disagreement at level-2 occurred when the annotators picked senses within the same level-1 class but different level-2 classes (e.g., CONTRAST versus CONCESSION). Similarly, disagreement at level-3 occurred when the annotators picked different senses under the same level-2 sense class (e.g., JUXTAPOSITION versus OPPOSITION).

Adjudication for disagreement at level-1 was done manually, by a team of experts, and disagreements at level-2 or level-3 were handled through automatic back-off to the next higher level. For example, a JUXTAPOSITION versus OPPOSITION disagreement would lead to a relation being automatically assigned their level-2 parent, namely, CONTRAST.

Annotation associated with automated back-off has contributed to there being only a level-1 sense annotation for 444 explicit and 257 implicit relations in the PDTB, almost all of which are either COMPARISON or EXPANSION. Although neither label is very informative, one might interpret such a label simply as under-specified with respect to its more specific level-2 daughters.

3.4.3 Multiple Concurrent Discourse Relations. Researchers using the PDTB for automated sense labeling of discourse relations have, by and large, assumed that its four level-1 senses (Figure 1) are disjoint. That is incorrect. Particular level-3 senses may be disjoint because they are defined as each other's inverse and hence can't both hold— for example, REASON and RESULT, PRECEDENCE and SUCCESSION, EXPECTATION and CONTRA-EXPECTATION. Other senses may be disjoint because their defining inferences contradict one another: for example, REASON requires **Arg2** to precede or coincide with *Arg1*, whereas PRECEDENCE requires that *Arg1* precede **Arg2**. They cannot both hold. However, most senses are compatible.

Table 3
Most common (≥10) pairs of level-2 sense labels on the 999 multiply-labeled explicit discourse relations. In pairs above the double line, one of the senses is temporal.

count	connective		senses
50	after	Contingency.Cause	Temporal.Asynchronous
30	and	Expansion.Conjunction	Temporal.Asynchronous
145	as	Contingency.Cause	Temporal.Synchrony
30	meanwhile	Comparison.Contrast	Temporal.Synchrony
92	meanwhile	Expansion.Conjunction	Temporal.Synchrony
10	since	Contingency.Cause	Temporal.Asynchronous
66	when	Contingency.Cause	Temporal.Asynchronous
41	when	Contingency.Cause	Temporal.Synchrony
65	when	Contingency.Condition	Temporal.Synchrony
12	when	Contingency.Condition	Temporal.Asynchronous
59	while	Comparison.Contrast	Temporal.Synchrony
21	while	Expansion.Conjunction	Temporal.Synchrony
138	and	Contingency.Cause	Expansion.Conjunction
13	but	Expansion.Conjunction	Comparison.Pragmatic contrast
10	if	Comparison.Concession	Contingency.Condition
11	while	Comparison.Contrast	Expansion.List

This is evident in the fact that annotators were allowed to assign up to two sense labels to each explicit or implicit connective, representing concurrent discourse relations. With explicit connectives, 999 of the 18,459 tokens (5.4%) were assigned two concurrent sense labels, with the most common pairs shown in Table 3.

Is 5.4% an accurate indicator of the frequency of multiple concurrent discourse relations between two arguments when they are linked by an explicit discourse connective? Evidence for a higher figure comes from an early experiment with two connectives, *since* and *when* (Miltsakaki et al. 2005). There, two annotators were given the option of labeling relations linked by one of these connectives in the WSJ corpus as either TEMPORAL or CAUSAL or TEMPORAL/CAUSAL, to indicate that both senses were conveyed. (The experiment was done on the 184 relations in the corpus headed by *since* and the first 100 relations headed by *when*, out of a total of 989.[6]) Those headed by *since* were annotated TEMPORAL/CAUSAL 21 times by one of the annotators (11.3%) and 16 times by the other (8.6%). Those headed by *when* were annotated TEMPORAL/CAUSAL even more frequently: 22% by one annotator and 28% by the other.

Compare this with the counts for multiply-labeled *since* and *when* in Table 3: Only 10/184 tokens of *since* (5.4%) were annotated with both a TEMPORAL and a CAUSAL sense, and only 184/989 tokens of *when* (18.6%) were annotated with both some kind of TEMPORAL and some kind of CAUSAL sense. In both cases, this is significantly less frequent than in the earlier experiment, suggesting that if annotators are not given explicit joint-sense options (such as TEMPORAL/CAUSAL or more specific pairs) and only invited to use multiple concurrent sense labels if they take multiple discourse relations to hold, their use of multiple labels may be intermittent at best. This is a loss to both language technology and theoretical and psycholinguistic understanding of discourse relations, and a situation that deserves to be fixed.

---

6 Miltsakaki et al. (2005) reported 186 tokens of *since* as discourse connectives; PDTB-Group (2008) subsequently reported 184 tokens. Most likely, two were later found not to be connectives.

In the case of implicit discourse relations, annotators could assign more than one sense label to a single implicit connective or they could insert more than one implicit connective, which were then individually sense-labeled. Both options indicated that concurrent discourse relations could be taken to hold between the specified arguments. Of the 16,224 implicit relations, 359 (2.2%) were annotated with a single implicit connective with multiple senses, and 171 (1.1%) were annotated with two implicit connectives, each taken to have a single sense. Both of these are very small numbers, so no hard conclusions can be drawn. Nevertheless, one might sample whether more of the implicit relations annotated with some CAUSAL sense might be more accurately annotated with some TEMPORAL sense as well.

We close by noting that of the 171 cases annotated with two implicit connectives, with each assigned a single sense, over half (93/171 = 54.4%) involved a connective paired with *for example*, *for instance*, or *for one thing* (e.g., *since, for example*; *as, for instance*; *because, for one thing*); 13 more (7.5%) were paired with *in particular* or *specifically* (e.g., *in particular, because*; *specifically, because*) and another 13 (7.5%) were paired with *in fact* (e.g., *although, in fact*; *so, in fact*). All but *in fact* are really connective modifiers (Section 3.1), even though they can also appear separately as connectives in their right. Such cases deserve further analysis, in connection with getting a better understanding of modified connectives, their prevalence, and their semantics.

4. Annotated Corpora Comparable to the PDTB

We noted in Section 1 that release of the PDTB has spawned similar efforts to annotate resources in other languages and genres following a lexically grounded approach to discourse relations. We also noted that these efforts vary in interesting ways from that of the PDTB. Here we describe both the nature and the sources of this variation, so that people contemplating development of comparable resources in additional languages and/or genres will recognize variation that is appropriate to their situation, while avoiding unnecessary variation that prevents inter-operability of these comparable corpora (Bunt, Prasad, and Joshi 2012).

Table 4 identifies the corpora we will discuss and the extent of their current annotation: the BioDRB (Prasad et al. 2011), the Leeds Arabic Discourse TreeBank, or LADTB (Al-Saif and Markert 2010, 2011; Al-Saif 2012), the Chinese Discourse TreeBank (Xue 2005; Zhou and Xue 2012; Zhou and Xue (in press)), the Turkish Discourse Bank or TDB (Zeyrek et al. 2008, 2009; Aktaş, Bozşahin, and Zeyrek 2010; Zeyrek et al. 2010; Demirsahin et al. 2013; Zeyrek et al. 2013), the Hindi Discourse Relation Bank (Oza et al. 2009; Kolachina et al. 2012; Sharma et al. 2013), and the Prague Discourse TreeBank, or PDiT (Mladová, Zikánová, and Hajičová 2008; Jínová, Mírovský, and Poláková 2012; Rysová 2012; Poláková et al. 2013), now part of the Prague Dependency TreeBank, version 3.0, PDT 3.0 (Bejček et al. 2013). (A comparable discourse treebank is being developed for French (Danlos et al. 2012), but it has not yet been released and the information needed to compare it to the other corpora in Table 4 is not available.)

Although these comparable corpora differ in ways to be discussed subsequently, they all adhere to the key ideas of PDTB annotation (Section 2) in being neutral to any discourse structure beyond the argument structure of individual discourse relations and in grounding discourse relations in lexical expressions. Where they annotate implicit discourse relations (Table 4), these comparable corpora follow the PDTB in annotating an inferred lexical grounding.

All of the corpora also follow the PDTB in taking discourse relations to hold between two and only two abstract objects, called Arg1 and Arg2, each associated with

Table 4
Comparison of the PDTB and comparably annotated corpora. *Count* is the number of annotated relations; *Coverage* is the text genre(s) in the corpus; *Mods*=**Y** if connective modifiers are annotated. *Impl*=**Y** if implicit connectives are annotated. *EntR*=**Y** if Entity Relations are annotated. *AltL*=**Y** if Alternative Lexicalizations are annotated. *Attr*=**Y** if attribution is annotated. *Supp*=**Y** if arguments can have supplementary text. *Sens*=**Y** if senses have been annotated. *Mult*=**Y** if multiple sense relations can be annotated for a single connective.

Name	Coverage	Count	Mods	Impl	EntR	AltL	Attr	Supp	Sens	Mult
PDTB	*WSJ* news, essays	40,600	Y	Y	Y	Y	Y	Y	Y	Y
BioDRB	Biomed papers	5,859	Y	Y	N	Y	N	N	Y	Y
LADTB	Arabic news	6,328	Y	N[1]	N	N	N	N	Y	Y
Chinese DTB	*Xinhua* news	3,951	Y	Y[2]	N	Y	N	N	Y	N
Turkish DB	novels, news, etc.	8,484	Y	N	N	N	N	Y	N	N
Hindi DRB	news	~5K	Y	Y	Y	Y	N	N	Y	Y
PDT 3.0 (PDiT 1.0)	news	20,542	Y	N	Y[3]	Y	N	N	Y	Y

[1]~70% of adjacent sentences in the LADTB are linked by an explicit connective, compared with ~12% in the PDTB.

[2]In 20 randomly selected files, over 80% of DRels were found to be implicit, compared with around 54.5% in the PDTB (Zhou and Xue 2012).

[3]Included in coreference annotation.

a possibly discontinuous text span. Although not every corpus annotates atttribution, where it is annotated, it is separate from the annotation of discourse relations.

4.1 Annotation Workflow

Because one purpose of this section is to inform people considering the development of similar resources in other languages and genres, we will briefly mention how workflow has varied in the development of comparable corpora and how it has affected annotator effort and inter-annotator agreement.

Workflow on the PDTB itself was based on the idea of using easier tasks to pave the way for more difficult ones. In practice, this meant separating the annotation of explicit and implicit relations, as explicits were perceived as easier to annotate. Explicit discourse relations were annotated one connective at a time throughout the corpus, before moving on to the next one on the list. The rationale for this was to improve the annotators' ability to annotate a particular connective by focusing their attention on that connective. Subsequently, implicit discourse relations were annotated document by document, analyzing each pair of adjacent sentences within each paragraph, as described in Section 3.2. Later, senses were annotated for explicit and implicit discourse relations—explicits by connective, and implicits by document.

Even for annotating explicit relations, this is not the only possible workflow. In annotating the LADTB (Al-Saif 2012; Al-Saif and Markert 2011), the nature of Modern Standard Arabic (MSA) demanded a different workflow. In MSA, as in English, words that can function as discourse connectives also have non-discourse functions. As such, confirming that a potential connective has a discourse function is directly related to identifying its arguments. One common form of argument to discourse relations in

Arabic news texts is an *Al-maSdar* noun, which is a tense-less expression of an event.[7] Their frequency affected annotation workflow in the LADTB.

The LADTB used a workflow for annotating explicit discourse relations that involved highlighting for the annotators all potential discourse connectives (including word-initial clitics), based on a pre-compiled list. An annotator first read through the entire text to achieve an overall understanding, before stepping through the highlighted items one by one. In order to tell if a potential connective has a discourse function, the annotator would see whether it had arguments, including strings interpretable as al-maSdar nouns. Workflow thus involved simultaneous confirmation of potential discourse connectives and identification of their arguments. After that, the annotator would add the one or more senses that a relation expresses. If a potential connective did not express a discourse function, the annotator would note it and go on to the next highlighted item.

Workflow for the BioDRB (Prasad et al. 2011) was designed to address the difficulty perceived in annotating inter-sentential relations in scientific text. On encountering a new sentence, the annotator had to first mark its inter-sentential relation(s) with the prior discourse, and only then annotate any intra-sentential relations within it. In this way, annotators were made to first attend to relations that were harder to pin down, as they progressed in their sequential reading and annotation of the text.

Ongoing annotation of the Chinese Discourse TreeBank (Zhou and Xue 2012) follows a fully sequential annotation strategy, largely for a language-specific reason—the customary writing style of Chinese, which often does not bother to distinguish the end of a sentence (marked with a *full stop*) from the end of a clause (marked with a *comma*). This has two major consequences: No rigid distinction can be made between inter- and intra-sentential connectives, and annotators must consider implicit relations both between full-stop delimited sentences and comma-delimited clauses. (The latter have not been annotated in the current PDTB.) Annotating implicit relations between comma-delimited clauses results in many more implicit relations. Zhou and Xue (2012) report a 18–82% split in their data between explicit and implicit discourse relations, compared with a 46–54% split in the PDTB. As such, having a separate task to cover 18% of the data was disfavored.

Although the particular style of annotation should have no effect on the content of annotation, it can affect inter-annotator agreement. To this end, researchers developing the Hindi DRB (Oza et al. 2009; Kolachina et al. 2012) carried out a systematic study of three workflow strategies (Sharma et al. 2013). The first strategy modeled the task exactly as in PDTB. In the second, explicits and implicits were annotated in exactly the order in which they were encountered on a sequential reading of the text. The third strategy operated per text, with annotators first marking all of its explicit connectives, and then its implicit relations, before moving on to another text. The latter two strategies were designed to ensure that annotators were aware of the coherence and flow of the discourse when carrying out the task. Sharma's findings show that better agreement is obtained when the annotators' attention is held to the text, but with no clear preference for a fully sequential approach (as in the second strategy), or an approach that separated the tasks on a text-by-text basis (as in the third).

Although ultimately, the choice of workflow may be language- or genre-specific, as noted for Chinese, the final choice, then, should be driven by considerations of

---

7 Although the PDTB admits nominalizations as arguments to explicit discourse connectives, they constitute only a small fraction of its arguments.

annotation reliability, which seems to be enhanced by *the annotators attending to the coherence and flow of the discourse*. Interoperability among these resources is not an issue here, as long as whatever strategy is used yields highly consistent annotation.

4.2 Inventory and Organization of Senses

The senses of discourse relations used in the PDTB and the hierarchy in which they are organized (Miltsakaki et al. 2008) drew on both in-house experiments (Miltsakaki et al. 2004, 2005) and previous work on the semantics of discourse relations (Lakoff 1971; Moens and Steedman 1988; Sweetser 1990; Jayez and Rossari 1998; Kehler 2002), among others. Neither has been adopted without some change in the comparable corpora: Additional senses have been introduced, while other senses have been eliminated or modified; the sense hierarchy has been modified, and in one case, abandoned. For example, Oza et al. (2009) propose a more general and uniform treatment of those discourse relations that are *pragmatic*, relating the speech act of one argument to either the content or speech act of the other. Al-Saif and Markert (2011) do the same. Almost every corpus includes at least one additional sense class, including SIMILARITY, PURPOSE, BACKGROUND, and GRADATION, among others, motivated more by the genre of the texts being annotated than by their language.

Changes in the PDTB sense hierarchy have been either at its root or its leaves: The BioDRB has eliminated the four top-level classes, adopting a two-level hierarchy. The Prague Discourse TreeBank (Poláková et al. 2012) now part of the PDT 3.0 and the LADTB (Al-Saif and Markert 2011) have also adopted a two-level hierarchy, but they preserve the top-level classes, collapsing the second and third levels of annotation. The Chinese Discourse TreeBank (Zhou and Xue 2012) has eliminated the hierarchy entirely, using a flat classification of just twelve sense categories. Sense annotation has not yet begun on the Turkish Discourse Bank.

From the standpoint of interoperability, a shared assumption about the meaning and classification of discourse relation senses is of utmost importance, because conflicts in the assumed meaning of labels would preclude any kind of comparative studies of the annotated resources, both within and across languages and domains. With the growing number of variations in sense annotation schema, we believe it is critical to collect the insights and findings from these studies and to find common threads, since we believe there is much that is common between them. Indeed, some recent work has usefully provided a mapping between their classification schemes and the PDTB classification (Prasad et al. 2011; Zhou and Xue 2012).

4.3 Annotation of Explicit Connectives

An obvious way in which the corpora vary is in the choice of explicit connectives to be annotated. Because of the rich morphology of Turkish, explicit connectives in Turkish include morphological suffixes attached to verb roots and complex subordinators consisting of a connective and a nominalizing suffix. The former have not yet been annotated in the Turkish Discourse Bank (TDB), although the latter have been (Zeyrek et al. 2013). Counterparts of the latter, called *phrasal expressions* in the TDB, appear as *AltLex* in the PDTB. As in the LADTB (Al-Saif 2012), nominalizations are commonly annotated as arguments in the TDB.

As well as being a morphologically rich language, with prefix clitics (as well as separate words and phrases) serving as explicit connectives, Arabic writing tends

towards long sentences conjoined with coordinating conjunctions (Ostler 1987), with the equivalent of *and* commonly found at the beginning of sentences and paragraphs (Al-Saif 2012). It was so common at the beginning of paragraphs in the newswire text annotated in the LADTB that all such tokens were simply assigned a CONJUNCTION relation to the closest proposition, unless a clearer discourse relation was explicitly indicated.

4.4 Lexical Grounding for Implicit Discourse Relations

The approach used in lexically grounding implicit discourse relations seems to be language-specific. For English, the PDTB's lexically grounded approach led to guidelines for annotating implicit relations in which annotators were asked to identify one or more connectives that could be inserted between proposed arguments to express the discourse relation(s) they took to hold between them (cf. Section 3.2). This was meant to serve as explicit evidence for their decisions.

For the Chinese Discourse TreeBank, Zhou and Xue (2012) adopt a different strategy, effectively using paraphrase rather than insertion. This is because, in a majority of cases,

> the wording rejects insertion of a connective even if it expresses the underlying discourse relation exactly (or sometimes, maybe the wording itself is the reason for not having a connective). (Zhou and Xue 2012, page 73)

This suggests that Chinese may use particular syntactic constructions to indicate intra-sentential discourse relations even more than English and German do (Meyer and Webber 2013). Thus, instead of having their annotators *insert* explicit connectives, Zhou and Xue (2012) have them *paraphrase* the relation between proposed arguments in terms of explicit connectives that typically express each discourse relation. These prototypical connectives then serve as the lexical grounding for the relation.

Although this is the only case we are aware of that has used a different approach to lexically grounding implicit relations, it is something that corpus developers should keep in mind, especially when considering the annotation of discourse relations within sentences.

4.5 Locus of Implicit Relations

The corpora differ in where they look for implicit discourse relations. As noted in Section 3.2, implicit relations were only considered in the PDTB between adjacent sentences within the same paragraph. Although a sentence might have an implicit relation to a sentence further afield, we decided that it would add too much to an already costly effort to have annotators seek them out. With respect to implicit discourse relations within a single sentence, we are aware of having deliberately ignored (for lack of resources) discourse relations that we know are there.

Where to look for implicit discourse relations is, in part, language-specific. We have already noted (Section 4.1) that the structure of Chinese sentences is such that a much larger proportion of discourse relations in Chinese occur intra-sententially. Hence the greater need to look for them there.

As for looking for discourse relations further afield, the comparable corpora vary, but not for language-specific reasons. Rather, it follows from the cost–coverage decision that all annotation efforts face. In this case, in both the Hindi Discourse Resource and the

BioDRB, implicit discourse relations have been sought more widely, allowing a sentence to be related to a non-adjacent sentence within the same paragraph.

4.6 Naming Convention for Arg1 and Arg2

One final way in which corpora comparable to the PDTB vary is with respect to the naming convention for *Arg1* and **Arg2**. In the PDTB, the choice follows syntactic criteria: With explicit discourse connectives, **Arg2** is the argument syntactically bound to the connective, and *Arg1* is the other argument. With implicit connectives, *Arg1* is the left-adjacent sentence, and **Arg2**, the right-adjacent one.

Although the same convention is followed in the BioDRB and Turkish Discourse Bank, the Chinese Discourse TreeBank, the Prague Discourse TreeBank, and the Hindi Discourse Resource have followed a semantically driven convention, in which arguments that play the same semantic role have the same label. This then eliminates level-3 senses in the PDTB sense hierarchy (e.g., reason/result, expectation/contra-expectation, precedence/succession) whose only purpose is to reflect the different linear order of the arguments. Again, while these differences are admissible without impacting the annotation scheme in any major way, comparative studies using these corpora need to be sensitive to these differences. We note, however, that in an experiment using this strategy for the Hindi annotation, Kolachina et al. (2012) report poor agreement for arguments of relations, and speculate that it was harder for annotators to use the semantic labeling convention.

4.7 Summary

Discourse annotation efforts that have followed the PDTB in adopting a lexically grounded (or adjacency-based) approach to annotation nevertheless differ from the PDTB in ways discussed earlier. Still, it appears that none of these differences is so great as to affect their interoperability with the PDTB or each other, or their use in multilingual language technology or machine translation (Meyer 2011; Meyer and Popescu-Belis 2012; Meyer and Webber 2013).

5. Complementary Annotations

Some of the linguistic phenomena annotated in the PDTB have also been annotated in connection with other levels of linguistic annotation—in particular, the temporal annotation of the *Wall Street Journal* portion of the Penn Treebank corpus found in TimeBank 1.2 (Pustejovsky et al. 2003a) and the verb-argument annotation found in PropBank (Palmer, Gildea, and Kingsbury 2005). Here we describe how these annotations are related.

We had both practical and theoretical motivation for carrying out the work described here. From a practical perspective, it might allow for future merging of annotation layers (Pustejovsky et al. 2005), future seeding of one annotation layer with another, and/or future consistency checking based on constraints between annotation at different levels. From a deeper, theoretical perspective, the work has the potential to lower, or even remove, barriers that have long existed between linguistic research at the sentence level and at the discourse level—barriers that have been equally obstructive to research in computational linguistics. This work can thus be seen as a small step towards "the transition from sentence to discourse."

5.1 TimeML and TimeBank

The TimeML temporal/event annotation (Pustejovsky et al. 2003b) of texts from the Penn TreeBank forms part of the TimeBank 1.2 corpus (Pustejovsky et al. 2003a).

TimeML supports the annotation of events, time periods, and temporal relations, expressed either explicitly or implicitly in a text. The information that TimeML makes explicit includes *temporal expressions* such as the date 10/26/1989 (tagged as TIMEX3), *temporal events* such as Nigel Lawson resigning as Chancellor of the Exchequer (tagged as EVENT), *temporal signals* such as *after*, *during*, and *in* (tagged as SIGNAL), and *temporal relations* between pairs of temporal expressions or event instances, or between a temporal expression and an event instance (tagged as TLINK). The set of *temporal relations* comes from Allen (1984). When a *temporal relation* is explicitly indicated by a *temporal signal*, the signal is included in the TLINK. This enables a clear correspondence with the PDTB.

A *temporal event* is annotated on the head of the syntactic construction that expresses it—the verb, in the case of a clause, as in the annotation of *resume* and *warrant* in Example (25), where the SIGNAL *until* is asserted to signal the temporal relation between *resume* and *warrant*.

(25)   He <EVENT>said</EVENT> <EVENT>construction</EVENT> wouldn't
         <EVENT>*resume*</EVENT> <SIGNAL>*until*</SIGNAL> market conditions
         <EVENT>*warrant*</EVENT> it. [wsj_0610]

This corresponds to the PDTB annotation:

(26)   He said *construction wouldn't resume* <u>until</u> (TEMPORAL.ASYNCHRONOUS.
         PRECEDENCE) **market conditions warrant it**.

But TimeML also annotates events expressed as nominalizations (e.g., *construction* in Example (25)) and simple nouns (e.g., *tax* in Example (27)).

(27)   And while there was no profit this year from discontinued operations,
         last year they <EVENT>contributed</EVENT> $34 million,
         <SIGNAL>before</SIGNAL><EVENT>tax</EVENT>. [wsj_0127]

Because *temporal events* are not limited to clauses, signals of *temporal relations* are not limited to clausal coordinators or subordinators or discourse adverbials, but also include prepositions such as *before* in Example (27). This is not annotated in the PDTB.

TimeML also allows for the annotation of certain non-temporal relations between events, including *conditional*, *evidential*, *non-evidential*, and *factive* relations. These are tagged SLINK (for Subordination Link). As with *temporal relations*, when these non-temporal relations are indicated with a SIGNAL (such as *if* for a *conditional* relation), the SIGNAL is included in the SLINK.

Although many of the same linguistic elements have been annotated in both the PDTB and TimeBank, the annotation itself can be quite different. The simplest difference relates to the sense of temporal relations: TimeML allows more specific relations between events than the PDTB's three broad relations TEMPORAL.ASYNCHRONOUS.PRECEDENCE (before), TEMPORAL.ASYNCHRONOUS.SUCCESSION (after), and TEMPORAL.SYNCHRONY (same time). For example, TimeML annotators can indicate that one event is *immediately before* or *immediately after* another, although TimeBank annotators have not used this when annotating relations signalled by *before* or *after*.

A more significant difference lies in where temporal relations are inferred in the PDTB and TimeBank. As noted earlier, the PDTB aims to annotate every discourse

relation—including temporal relations—that holds between abstract objects (mainly clausal or sentential interpretations) that are signaled by an explicit discourse connective (or some *alternative lexicalization* of a connective) or by the fact of sentence adjacency. In the latter case, either an implicit discourse relation will be inferred between them, or a relation involving some entity mentioned in the first sentence (*EntRel*), or no relation at all (*NoRel*). In all cases, something will be annotated.

In contrast, TimeML guidelines specify that if a temporal relation is explicitly signaled in the text, then events and/or time periods specified in different sentences may be linked through signals such as *previously*, *earlier*, *at the same time*, *then*, or *meanwhile*. If no temporal relation is explicitly signaled, then temporal elements in different sentences are not linked, so there are no TLINKS in TimeBank corresponding to the PDTB's 950 implicit temporal relations between adjacent sentences.

On the other hand, the TimeML guidelines allow a temporal relation to be inferred from discourse relations that are not primarily temporal. For example, some discourse relations annotated in the PDTB as causal (i.e., CONTINGENCY.CAUSE.REASON or CONTINGENCY.CAUSE.RESULT) are annotated as temporal relations in TimeBank because both arguments express temporal events and because a cause event starts before its result. This is the case in Example (28), where TimeBank annotates the *holding* event as occurring before the *adjusting* event, which has a negative polarity attribute.[8] In the PDTB, only the explicitly signaled causal relation is annotated.

(28)   *Previously, Columbia didn't have to adjust the book value of its junk-bond holdings to reflect declines in market prices*, <u>because</u> (CONTINGENCY.CAUSE.REASON) **it held the bonds as long-term investments**. [wsj_1013]

However, not all relations annotated in the PDTB as CONTINGENCY.CAUSE.REASON have a corresponding temporal annotation in TimeBank: Those corresponding to generic statements (e.g., Example (29)) do not, because generic statements are not taken to express temporal relations.

(29)   *It's harder to sell stocks when the sell programs come in* <u>because</u> (CONTINGENCY. CAUSE.REASON) **some market makers don't want to {take the orders}**. [wsj_0585]

Finally, there is some correspondence between PDTB annotation and the non-temporal relations that TimeBank annotates as SLINK. What TimeBank annotates as a *conditional* SLINK overlaps with the explicit discourse relations annotated in the PDTB as CONTINGENCY.CONDITION. Other types of SLINK (*modal*, *factive*, *evidential*, and *negative evidential*) are related to the properties of *Attribution* in the PDTB (cf. Section 3.3; Prasad et al. 2007).

Enriching TimeBank relations based on annotation in the PDTB, and vice versa, would require a more detailed study of both the annotation frameworks and annotation practice in the two corpora. The same would go for using the annotation in one as a consistency check on the other. Nevertheless, both would potentially be of great value to delivering more usefully annotated resources.

---

8 http://www.timeml.org/site/publications/timeMLdocs/annguide_1.2.1.pdf, page 44.

```
a. ( (S
     (SBAR-TMP:ARGM-TMP (IN Until)
       (S
         (NP-SBJ-1 (DT the) (NN building) )
         (VP (VBZ is)
           (VP (VBN completed)
             (NP (-NONE- *-1) )))))
     (, ,)
     (NP-SBJ:ARG0 (NNP Exxon) )
     (VP (MD:ARGM-MOD will)
       (VP (VB rent)
         (NP:ARG1
           (NP (NN part) )
           (PP (IN of)
             (NP (DT an) (VBG existing) (NN office) (NN tower) )))))
     (. .) ))
```

b. <u>Until</u> (TEMPORAL.ASYNCHRONOUS.PRECEDENCE) *the building is completed*, **Exxon will rent part of an existing office tower**. [wsj_0784]

Figure 2
(a) PropBank annotation of the verb *rent*; (b) PDTB annotation of the sentence that *rent* heads.

5.2 PropBank

More interesting is the relation between PDTB annotation of discourse relations and PropBank annotation (Palmer, Gildea, and Kingsbury 2005) of the sentence-internal argument structure. PropBank provides, for each verb predicate in the Penn TreeBank, its sense and the semantic role of each of its arguments. An argument can be either required by the verb through its valency and assigned an index such as Arg0, Arg1, and so on, or accepted as a modifier (annotated with an ArgM label). ArgM arguments are further assigned functional tags such as MNR (manner), MOD (modal), TMP (temporal), CAU (causal), DIS (discourse), and so forth. For example, the PropBank annotation of one instance of the verb *rent* is shown in Figure 2a. Besides its subject and object (Arg0 and Arg1), the modal auxiliary *will* is annotated as ArgM-MOD and the subordinate clause headed by *until* is annotated ArgM-TMP.

Many of the ArgM arguments in PropBank are either clauses or nominalizations that denote events. Many of these align with discourse relations in the PDTB. For example, the ArgM annotation of the subordinate clause in Figure 2a corresponds to **Arg2** of the PDTB annotation of the discourse relation associated with *until* (Figure 2b).

We can quantify the extent and nature of this correspondence between the two annotation layers, and in doing so consider two related questions: (1) How many and to what extent are the intra-sentential relations in PDTB also accounted for by the dependencies annotated in PropBank, and (2) Are there gaps in the discourse-level annotation that can be identified from the PropBank layer? Our analysis here is in terms of the annotation in PropBank-1.[9] Although the scope of this annotation has been extended within the Ontonotes project (Weischedel et al. 2012), this does not affect our general points.

5.2.1 Correspondence of PDTB Intra-sentential Relations with PropBank. We first assess whether the intra-sentential relations in the PDTB can be fully accounted for by the verb-ArgM dependencies in PropBank in terms of quantity, content, and consistency. If so, annotating them again at the discourse level would have involved needless repetition.

---

9 `http://catalog.ldc.upenn.edu/LDC2004T14.`

942

There are 11,830 intra-sentential relations annotated in the PDTB, accounting for 28% of the discourse relations annotated in the corpus. Of these 11,830 relations (all of which are candidates for overlap with PropBank), 11,236 involve explicit connectives and 594 do not. The latter primarily hold between independent clauses separated by a semicolon. As the clauses so-linked are independent, neither being a modifier of the other, these relations are not covered in PropBank.

The other 11,236 intra-sentential relations include relations between clauses linked by an explicit coordinating conjunction (such as *and* and *but*). Like the semicolon-separated clauses, these are independent and so also outside the scope of PropBank. The set also includes relations between clauses in the same sentence signaled by a discourse adverbial. In PropBank, discourse adverbials are generally annotated as discourse-linking modifiers (ARGM-DIS), leaving unspecified what they link to. For example, in Example (30), the discourse adverbial *instead* conveys a relation between the two "throwing" propositions. While PropBank annotates *instead* as ARGM-DIS of the main clause *throw*, it does not explicitly link *instead* to its other argument.

(30)   When the champ has lost his stuff, the great mystery novelist wrote, *when he can no longer throw the high hard one*, **he throws his heart** <u>instead</u>. [wsj_1649]

PropBank-1 did not annotate arguments to copula verbs, so subordinating clauses attached to these verbs were not covered. However, copula verbs have subsequently been included in extensions covering over 75% of PropBank-1 and released as part of Ontonotes-5.0, so these subordinating clauses are now marked as arguments.

Besides differing in terms of intra-sentential coverage, PropBank and the PDTB also differ in their semantics. Specifically, even those PropBank ArgM roles that are closest to discourse relations—ArgM-CAU (causal), ArgM-TMP (temporal), ArgM-PNC (purpose), ArgM-MNR (manner), and ArgM-ADV (adverbial)—differ from the semantics of PDTB senses in several ways.

- Specificity: ARG-TMP is annotated where the PDTB annotates a more specific sense (Synchrony, Precedence, and Succession) of its top-level class TEMPORAL.

- Heterogeneity: Subordinate clauses that are labeled as ArgM-ADV correspond to the full range of PDTB senses.

- Multiplicity: The PDTB allows more than one sense label to be associated with a single discourse connective to indicate that multiple sense relations hold concurrently (e.g., a token of *since* may be labeled with both a temporal and causal sense). In contrast, PropBank only permits a constituent to fill a single functional role. Nevertheless, the seven cases where subordinate clauses are annotated as causal (ArgM-CAU) in PropBank and some form of TEMPORAL sense in the PDTB reveal additional cases of the earlier-mentioned under-annotation of multiple concurrent senses in the PDTB (Section 3.4.3).

- Coverage: The PDTB's sense inventory currently lacks senses corresponding to PropBank's ArgM-MNR and ArgM-PNC roles. This will be discussed in Section 5.2.2.

In terms of consistency, there are also some mismatches in alignment between PDTB arguments and PropBank's semantic role structure, due to the fact that PropBank

annotation is tied directly to the syntactic trees in the PTB. Figure 3(b) shows the PropBank annotation of the verb *say* over the PTB parse tree shown in Figure 3(a), with its initial *when*-clause parsed as a temporal modifier (ArgM-TMP) of *say*. In contrast, PDTB annotation has been done over the raw text, with discontinuous spans permitted as arguments. This allows attribution to be included or excluded from a discourse relation (Section 3.3). In this case, Figure 3(c) shows attribution is excluded: The temporal relation (SUCCESSION) is annotated between *winning* and *awarding*, implying that Mr. Green's winning of the verdict was followed by the judge giving him the additional award. Given the difference in annotation practice, the extent of such mismatches between PDTB and PropBank is expected to be the same as that between PDTB and PTB (Dinesh et al. 2005).

In this section, we have considered whether the intra-sentential relations in the PDTB (i.e., those with both arguments in the same sentence) can be fully accounted for by the verb-ArgM dependencies in PropBank. We have shown that the account is only partial, due in part to the significantly different goals of the two annotation projects and in part to differences in methodological choices. As such, even intra-sententially, a separate layer of discourse relation annotation is motivated.

```
(a) ( (S
     (SBAR-TMP
       (WHADVP-1 (WRB When) )
       (S
         (NP-SBJ (NNP Mr.) (NNP Green) )
         (VP (VBD won)
           (NP
             (NP (DT a)
               (ADJP ($ $) (CD 240,000) (-NONE- *U*) )
               (NN verdict) )
             (PP (IN in)
               (NP
                 (NP (DT a) (NN land) (NN condemnation) (NN case) )
                 (PP (IN against)
                   (NP (DT the) (NN state) )))))
           (ADVP-TMP (-NONE- *T*-1) )
           (PP-TMP (IN in)
             (NP (NNP June) (CD 1983) )))))
     (, ,)
     (NP-SBJ (PRP he) )
     (VP (VBZ says)
       (SBAR (-NONE- 0)
         (S
           (NP-SBJ (NNP Judge) (NNP O'Kicki) )
           (VP
             (ADVP (RB unexpectedly) )
             (VBD awarded)
             (NP (PRP him) )
             (NP (DT an) (JJ additional) ($ $) (CD 100,000) (-NONE- *U*) )))))
     (. .) ))
```

(b) **rel=** *say*;   **Arg0=** *he*;
Arg1="Judge O'Kicki unexpectedly awarded him an additional $100,000.";
ArgM-TMP="When Mr. Green won a $240,000 verdict in a land condemnation case against the state in June 1983"

(c) <u>When</u> (TEMPORAL.ASYNCHRONOUS.SUCCESSION) **Mr. Green won a $240,000 verdict in a land condemnation case against the state in June 1983**, he says *Judge O'Kicki unexpectedly awarded him an additional $100,000*. [wsj_0267]

Figure 3
Comparison of (a) Penn TreeBank; (b) PropBank; (c) PDTB annotation.

5.2.2 Potential of Seeding New Discourse Relations from PropBank. Next, we assess whether any of the verb-ArgM dependencies in PropBank could potentially correspond to discourse relations that are not yet annotated in the PDTB. If the type and number of such relations is significant, then PropBank annotations could be used to seed the PDTB with new relations, which could then be corrected and/or annotated manually.

To do this, we aligned the PropBank annotations with the PTB and the PDTB. We started by considering only the five ArgM types mentioned earlier, which gave us a total of 43,432 verb–ArgM dependencies. For ease of analysis, we ignored the tokens of split ArgMs (i.e., ArgMs that are not spanned by a single node). We also ignored tokens from WSJ texts that were not included in the PDTB distribution because of problems with conversion of the parsed files to stand off annotation format (PDTB-Group 2008).

From the total of 43,432 ArgMs, we identified 11,538 ArgMs as clausal, using the PropBank alignment with the PTB. 4116 of these are *free adjuncts* (Example (31)), all of which are new potential discourse relations to consider for the PDTB (cf. Section 3.2). The remaining 7,422 clauses either start with a subordinator or subordinating conjunction, including both finite (Example (32)) and non-finite (Example (33)) adverbial clauses, or are reduced clauses (Example (34)). For these explicitly subordinated clauses, PropBank alignment with the PDTB shows that explicit subordinators/subordinating conjunctions were annotated as connectives in 5,471 of the 7,422 ArgMs, leaving the remaining 1,951 ArgMs as new potential discourse relations for the PDTB. Each of these 6,067 new potential relations identified from the PropBank (i,e., the 4,116 free adjuncts and the 1,951 subordinated clauses) would still have to be reviewed manually to determine whether it does in fact fulfill a discourse function or not.

(31)   They say greedy market manipulators have *made* a shambles of the nation's free-enterprise system [$_{ArgM-ADV}$ turning the stock market into a big gambling den, with the odds heavily stacked against the small investor].

(32)   That $130 million, Mr. Sherwood said, "*gives* us some flexibility [$_{ArgM-CAU}$ in case Temple raises its bid].

(33)   Those dividend bulls argue that corporations are in the unusual position of *having* plenty of cash left over [$_{ArgM-TMP}$ after paying dividends and making capital expenditures].

(34)   [$_{ArgM-ADV}$ If not for a 59.6% surge in orders for capital goods by defense contractors], factory orders would have *fallen* 2.1%.

New potential relations identified through PropBank would allow for expanding not just the number of PDTB relations, but also the repertoire of connectives (such as *in case* in Example (32)) and sense categories, in particular a MANNER relation, corresponding to ArgM-MNR, and a PURPOSE relation, corresponding to ArgM-PNC, that are not currently covered in the PDTB. It is important to note here that manner and purpose arguments annotated in PropBank will only be considered arguments to discourse relations when they denote events, facts, states, or propositions, since these are what are taken to be arguments to discourse relations in the PDTB.

5.3 Summary

Linguistic annotation invariably involves a considerable amount of time and effort. When linguistic analysis at multiple levels is encoded on the same source corpus as different layers of annotation, there is potential value in assessing how the annotation

content of the different layers differ from each other, and in exploring how annotations from one layer can be exploited usefully for annotation in other layers. This section has compared the annotation content of the PDTB with that of TimeBank and PropBank, showing that while some of the linguistic phenomena annotated in PDTB have also been annotated in TimeBank and PropBank, there are significant differences in both the extent and the content of the annotation. This section has also discussed some of the ways in which annotations from one layer can enrich and/or improve the consistency of annotation in other layers.

6. Conclusion

Our goals in this paper have been to

- give a thoughtful description of the PDTB that reflects what we have learned since release of the corpus in 2008;

- correct some assumptions about the PDTB that show that researchers may either be ignoring significant features of its annotation of discourse relations or taking *accidental* properties of its annotation to be *intrinsic* properties of the discourse relations themselves;

- describe and place in context the ways in which annotation of comparable resources in other languages and genres has varied from that of the PDTB; and

- provide an analysis of the relation of PDTB annotation to that of TimeBank and PropBank over the same Penn TreeBank corpus and show how they are, in large part, complementary.

We hope that the number of researchers able to make use of the PDTB will continue to grow, as will the number of similarly annotated corpora. We ourselves hope to be able to enrich the PDTB in the future—widening the scope of discourse relations that are annotated, improving the recording of evidence for annotation decisions, and expanding the annotation to include additional textual genres, especially ones that are less formal than news texts, such as public talks and consumer health advice. The results should be of further benefit to a growing community of scholars and developers considering the challenges of extended text.

References

Agarwal, M., R. Shah, and P. Mannem. 2011. Automatic question generation using discourse cues. In *Proceedings of the ACL HLT 2011 Workshop on Innovative Use of NLP for Building Educational Applications*, pages 1–9, Portland, OR.

Aijmer, K. and A.-M. Simon-Vandenbergen. 2004. A model and a methodology for the study of pragmatic markers. *Journal of Pragmatics*, 36:1781–1805.

Aktaş, B., C. Bozşahin, and D. Zeyrek. 2010. Discourse relation configurations in Turkish and an annotation environment. In *Proceedings of the 4th Linguistic Annotation Workshop*, pages 202–206, Uppsala.

Al-Saif, A. 2012. *Human and automatic annotation of discourse relations for Arabic*. Ph.D. thesis, University of Leeds.

Al-Saif, A. and K. Markert. 2010. The Leeds Arabic Discourse Treebank: Annotating discourse connectives for Arabic. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC-2010)*, pages 2,046–2,053, Valletta.

Al-Saif, A. and K. Markert. 2011. Modelling discourse relations for Arabic. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 736–747, Edinburgh.

Allen, J. 1984. Towards a general theory of action and time. *Artificial Intelligence*, 23:123–154.

Asher, N. 1993. *Reference to Abstract Objects*. Kluwer, Dordrecht.

Asher, N. and A. Lascarides. 2003. *Logics of conversation*. Cambridge University Press.

Asr, F. T. and V. Demberg. 2012a. Implicitness of discourse relations. In *Proceedings of COLING*, pages 2,669–2,684, Mumbai.

Asr, F. T. and V. Demberg. 2012b. Measuring the strength of linguistic cues for discourse relations. In *Proceedings of the Workshop on Advances in Discourse Analysis and its Computational Aspects (ADACA)*, pages 33–42, Mumbai.

Asr, F. T. and V. Demberg. 2013. On the information conveyed by discourse markers. In *Proceedings of the 4th Annual Workshop on Cognitive Modeling and Computational Linguistics (CMCL)*, pages 84–93, Sofia.

Baldridge, J., N. Asher, and J. Hunter. 2007. Annotation for and robust parsing of discourse structure on unrestricted texts. *Zeitschrift fur Sprachwissenschaft*, 26:213–239.

Bejček, E., E. Hajičová, J. Hajič, P. Jínová, V. Kettnerová, V. Kolářová, M. Mikulová, J. Mírovský, A. Nedoluzhko, J. Panevová, L. Poláková, M. Ševčíková, J. Štěpánek, and Šárka Zikánová. 2013. Prague Dependency Treebank 3.0 data/software. Technical report, Univerzita Karlova v Praze, MFF, FAL, Prague. http://ufal.mff.cuni.cz/pdt3.0/.

Bunt, H., R. Prasad, and A. Joshi. 2012. First steps towards an ISO standard for annotating discourse relations. In *Proceedings of the Joint ISA-7, SRSL-3, and I2MRT Workshop on Semantic Annotation and the Integration and Interoperability of Multimodal Resources and Tools*, pages 60–69, Istanbul.

Carlson, L., D. Marcu, and M. E. Okurowski. 2001. Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Proceedings of the 2nd SIGDIAL Workshop on Discourse and Dialogue, Eurospeech 2001*, pages 1–10, Aalborg.

Danlos, L., D. Antolinos-Basso, C. Braud, and C. Roze. 2012. Vers le FDTB: French Discourse Tree Bank. In *Proceedings of the Joint Conference JEP-TALN-RECITAL*, pages 471–479, Grenoble.

Demirsahin, I., A. Ozturel, C. Bozsahin, and D. Zeyrek. 2013. Applicative structures and immediate discourse in the Turkish Discourse Bank. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 122–130, Sofia.

Dinesh, N., A. Lee, E. Miltsakaki, R. Prasad, A. Joshi, and B. Webber. 2005. Attribution and the (non)-alignment of syntactic and discourse arguments of connectives. In *Proceedings of the ACL Workshop on Frontiers in Corpus Annotation II: Pie in the Sky*, pages 29–36, Ann Arbor, MI.

Elwell, R. and J. Baldridge. 2008. Discourse connective argument identification with connective specific rankers. In *Proceedings of ICSC-2008*, pages 198–205, Santa Clara, CA.

Forbes-Riley, K., B. Webber, and A. Joshi. 2006. Computing discourse semantics: The predicate-argument semantics of discourse connectives in D-LTAG. *Journal of Semantics*, 23:55–106.

Ghosh, S., R. Johansson, G. Riccardi, and S. Tonelli. 2011a. Shallow discourse parsing with conditional random fields. In *Proceedings of the International Joint Conference on Natural Language Processing*, pages 1,071–1,079, Chiang Mai.

Ghosh, S., R. Johansson, G. Riccardi, and S. Tonelli. 2012. Improving the recall of a discourse parser by constraint-based postprocessing. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, pages 2,791–2,794, Istanbul.

Ghosh, S., S. Tonelli, G. Riccardi, and R. Johansson. 2011b. End-to-end discourse parser evaluation. In *Proceedings of the Fifth IEEE International Conference on Semantic Computing*, pages 169–172, Palo Alto, CA.

Halliday, M. A. K. and R. Hasan. 1976. *Cohesion in English*. Longman, London.

Hirschberg, J. and D. Litman. 1993. Empirical studies on the disambiguation of cue phrases. *Computational Linguistics*, 19(3):501–530.

Jayez, J. and C. Rossari. 1998. Pragmatic connectives as predicates: the case of inferential connectives. In P. St Dizier, ed., *Predicative forms in natural language and in lexical knowledge bases*, pages 285–319. Springer, Dordrecht.

Jiang, X. 2013. *Predicting the use and interpretation of implicit and explicit discourse connectives*. Ph.D. thesis, M.Sc. Thesis, School of Psychology, Philosophy and Language Sciences (PPLS), University of Edinburgh.

Jínová, P., J. Mírovský, and L. Poláková. 2012. Semi-automatic annotation of intra-sentential discourse relations in PDT. In *Proceedings of the Workshop on Advances in Discourse Analysis and its Computational Aspects (ADACA)*, pages 43–58, Mumbai.

Kehler, A. 2002. *Coherence, Reference, and the Theory of Grammar*. CSLI Publications, Palo Alto, CA.

Knott, A. 1996. *A Data-Driven Methodology for Motivating a Set of Coherence Relations*. Ph.D. thesis, University of Edinburgh.

Knott, A., J. Oberlander, M. O'Donnell, and C. Mellish. 2001. Beyond elaboration: The interaction of relations and focus in coherent text. In T. Sanders, J. Schilperoord, and W. Spooren, editors, *Text Representation: Linguistic and Psycholinguistic Aspects*, pages 181–196. John Benjamins Publishing.

Kolachina, S., R. Prasad, D. M. Sharma, and A. Joshi. 2012. Evaluation of discourse relation annotation in the Hindi Discourse Relation Bank. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, pages 823–828, Istanbul.

Lakoff, R. 1971. Ifs, ands and buts about conjunction. *Studies in Linguistic Semantics*, 3:114–149.

Lin, Z., H. T. Ng, and M.-Y. Kan. 2012. A PDTB-styled end-to-end discourse parser. *Natural Language Engineering*, 20:151–184.

Mann, W. C. and S. A. Thompson. 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.

Marcus, M. P., B. Santorini, and M. A. Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313–330.

Martin, J. R. 1992. *English Text: System and Structure*. Benjamins, Amsterdam.

Meyer, T. 2011. Disambiguating temporal-contrastive connectives for machine translation. In *Proceedings of the ACL 2011 Student Session*, pages 46–51, Portland, OR.

Meyer, T. and A. Popescu-Belis. 2012. Using sense-labeled discourse connectives for statistical machine translation. In *Proceedings of the Workshop on Hybrid Approaches to Machine Translation (HyTra)*, pages 129–138, Avignon.

Meyer, T. and B. Webber. 2013. Implicitation of discourse connectives in (machine) translation. In *Proceedings of the ACL Workshop on Discourse in Machine Translation*, pages 19–26, Sofia.

Miltsakaki, E., N. Dinesh, R. Prasad, A. Joshi, and B. Webber. 2005. Experiments on sense annotation and sense disambiguation of discourse connectives. In *Proceedings of the Fourth Workshop on Treebanks and Linguistic Theories (TLT)*, pages 1–12, Barcelona.

Miltsakaki, E., R. Prasad, A. Joshi, and B. Webber. 2004. Annotating discourse connectives and their arguments. In *Proceedings of the Workshop on Frontiers in Corpus Annotation (Human Language Technology Conference and the Conference of the North American Association of Computational Linguistics)*, pages 9–16, Boston, MA.

Miltsakaki, E., L. Robaldo, A. Lee, and A. Joshi. 2008. Sense annotation in the Penn Discourse Treebank. *Computational Linguistics and Intelligent Text Processing, Lecture Notes in Computer Science*, 4919:275–286.

Mladová, L., Šárka Zikánová, and E. Hajičová. 2008. From sentence to discourse: Building an annotation scheme for discourse based on Prague Dependency Treebank. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, pages 2,564–2,570, Marrakech.

Moens, M. and M. Steedman. 1988. Temporal ontology and temporal reference. *Computational Linguistics*, 14(2):15–28.

Ostler, S. 1987. Academic and ethnic background as factors affecting writing

performance. In A. Purves, editor, *Writing across Languages and Cultures: Issues in Contrastive Rhetoric*, pages 261–272. Sage.

Oza, U., R. Prasad, S. Kolachina, S. Meena, D. M. Sharma, and A. Joshi. 2009. Experiments with annotating discourse relations in the Hindi Discourse Relation Bank. In *Proceedings of the 7th International Conference on Natural Language Processing (ICON)*, pages 1–10, Hyderabad.

Palmer, M., D. Gildea, and P. Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

Pareti, S. 2012. A database of attribution relations. In *Proceedings of the 8th Conference on International Language Resources and Evaluation (LREC12)*, pages 3,213–3,217, Istanbul.

Patterson, G. and A. Kehler. 2013. Predicting the presence of discourse connectives. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 914–923, Seattle, WA.

PDTB-Group. 2008. The Penn Discourse TreeBank 2.0 Annotation Manual. Technical report IRCS-08-01, Institute for Research in Cognitive Science, University of Pennsylvania.

Pitler, E. and A. Nenkova. 2009. Using syntax to disambiguate explicit discourse connectives in text. In *Proceedings of the Joint Conference of the 47th Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing*, pages 13–16, Singapore.

Pitler, E., M. Raghupathy, H. Mehta, A. Nenkova, A. Lee, and A. Joshi. 2008. Easily identifiable discourse relations. In *Proceedings of COLING*, pages 87–90, Manchester.

Poláková, L., P. Jínová, Šárka Zikánová, Z. Bedřichová, J. Mírovský, M. Rysová, J. Zdeňková, V. Pavlíková, and E. Hajičová. 2012. Manual for annotation of discourse relations in the Prague Dependency Treebank. Technical report TR-2012/47, Institute of Formal and Applied Linguistics, Charles University in Prague, Prague, Czech Republic.

Poláková, L., J. Mírovský, A. Nedoluzhko, P. Jínová, V. Zikánová, and E. Hajičová. 2013. Introducing the Prague Discourse Treebank 1.0. In *Proceedings of the 6th International Joint Conference on Natural Language Processing*, pages 91–99, Nagoya.

Polanyi, L., C. Culy, M. van den Berg, G. L. Thione, and D. Ahn. 2004. Sentential structure and discourse parsing. In *ACL 2004 Workshop on Discourse Annotation*, pages 80–87, Barcelona.

Prasad, R., N. Dinesh, A. Lee, A. Joshi, and B. Webber. 2007. Attribution and its annotation in the Penn Discourse TreeBank. *Traitement Automatique des Langues, Special Issue on Computational Approaches to Document and Discourse*, 47(2):43–64.

Prasad, R., N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, and B. Webber. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings of LREC*, pages 2,961–2,968, Marrakesh.

Prasad, R. and A. Joshi. 2008. A discourse-based approach to generating why-questions from texts. In *Proceedings of the Workshop on the Question Generation Shared Task and Evaluation Challenge*, pages 1–3, Arlington, VA.

Prasad, R., A. Joshi, and B. Webber. 2010a. Exploiting scope for shallow discourse parsing. In *Proceedings of the Seventh International Conference on Language Resources and their Evaluation*, pages 2,076–2,083, Valletta.

Prasad, R., A. Joshi, and B. Webber. 2010b. Realization of discourse relations by other means: Alternative lexicalizations. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 1,023–1,031, Beijing.

Prasad, R., S. McRoy, N. Frid, A. Joshi, and H. Yu. 2011. The Biomedical Discourse Relation Bank. *BMC Bioinformatics*, 12(188):1–18.

Pustejovsky, J., P. Hanks, R. Sauri, A. See, R. Gaizauskas, A. Setzer, and D. Radev. 2003a. The Timebank corpus. In *Proceedings of the Corpus Linguistics Meeting*, pages 647–656, Lancaster.

Pustejovsky, J., A. Meyers, M. Palmer, and M. Poesio. 2005. Merging PropBank, NomBank, TimeBank, Penn Discourse Treebank and Coreference. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*, pages 5–12, Ann Arbor, MI.

Pustejovsky, J., J. Castaño, R. Ingria, R. Sauri, R. Gaizauskas, A. Setzer, and G. Katz. 2003b. TimeML: Robust specification of event and temporal expressions in text. *New Directions in Question Answering*, 3:28–34.

Ramesh, B., R. Prasad, T. Miller, B. Harrington, and H. Yu. 2012. Automatic discourse connective detection in biomedical text. *Journal of the American*

Medical Informatics Association,
19(5):800–808.

Rysová, M. 2012. Alternative lexicalizations
of discourse connectives in Czech.
In *Proceedings of the 8th International
Conference on Language Resources and
Evaluation*, pages 2,800–2,807, Istanbul.

Sharma, H., P. Dakwale, D. Sharma,
R. Prasad, and A. Joshi. 2013. Assessment
of different workflow strategies for
annotating discourse relations: A case
study with HDRB. In A. Gelbukh, editor,
*Computational Linguistics and Intelligent Text
Processing, LNCS 7816*, pages 523–532.
Springer.

Stede, M. 2008. RST revisited: Disentangling
nuclearity. In C. Fabricius-Hansen and
W. Ramm, editors, *'Subordination' versus
'coordination' in sentence and text–from a
cross-linguistic perspective*, pages 33–58.
John Benjamins, Amsterdam.

Stede, M. 2012. *Discourse Processing*. Morgan
& Claypool Publishers.

Sweetser, E. 1990. *From Etymology to
Pragmatics: Metaphorical and Cultural
Aspects of Semantics*. Cambridge
University Press.

Versley, Y. 2010. Discovery of ambiguous and
unambiguous discourse connectives via
annotation projection. In *Proceedings of the
Workshop on the Annotation and Exploitation
of Parallel Corpora (AEPC)*, pages 83–92,
Tartu.

Webber, B. 2013. What excludes an
alternative in coherence relations? In
*Proceedings of the 10th International
Conference on Computational Semantics
(IWCS 2013)*, pages 276–287, Potsdam.

Webber, B., M. Egg, and V. Kordoni. 2012.
Discourse structure and language
technology. *Natural Language Engineering*,
18(4):437–490.

Weischedel, R., M. Palmer, M. Marcus,
E. Hovy, S. Pradhan, L. Ramshaw, N. Xue,
A. Taylor, J. Kaufman, M. Franchini,
M. El-Bachouti, R. Belvin, and A. Houston.
2012. Ontonotes release 5.0. Technical
report, Linguistic Data Consortium.

Wellner, B. 2009. *Sequence Models and
Re-ranking Methods for Discourse Parsing*.
Ph.D. thesis, Brandeis University,
Boston, MA.

Wellner, B. and J. Pustejovsky. 2007.
Automatically identifiying the arguments
of discourse connectives. In *Proceedings of
EMNLP-CoNLL*, pages 92–101.

Xue, N. 2005. Annotating discourse
connectives in the Chinese Treebank.
In *Proceedings of the ACL Workshop on
Frontiers in Corpus Annotation II: Pie in
the Sky*, pages 84–91, Ann Arbor, MI.

Zeyrek, D., Ümit Deniz Turan, and
I. Demirşahin. 2008. Structural and
presuppositional connectives in Turkish.
In A. Benz, P. Kühnlein, and M. Stede,
editors, *Constraints in Discourse 3*,
pages 131–137. University of Potsdam,
Germany.

Zeyrek, D., Ümit Deniz Turan, C. Bozşahin,
R. Çakıcı, A. Sevdik-Çallı, I. Demirşahin,
B. Aktaş, İhsan Yalçınkaya, and H. Ögel.
2009. Annotating subordinators in the
Turkish Discourse Bank. In *Proceedings of
the Third Linguistic Annotation Workshop
(LAW III), ACL-IJCNLP-2009*, pages 44–48,
Singapore.

Zeyrek, D., I. Demirşahin, A. Sevdik-Çallı,
H. Ögel, İhsan Yalçınkaya, and Ümit
Deniz Turan. 2010. The annotation scheme
of the Turkish Discourse Bank and an
evaluation of inconsistent annotations.
In *Proceedings of the Fourth Linguistic
Annotation Workshop (LAW-IV), ACL 2010*,
pages 282–289, Uppsala.

Zeyrek, D., I. Demirşahin, A. Sevdik-Çallı,
and R. Çakıcı. 2013. Turkish Discourse
Bank: Porting a discourse annotation
style to a morphologically rich language.
Dialogue and Discourse, 4(2):174–184.

Zhou, Y. and N. Xue. (in press). The Chinese
Discourse TreeBank: A Chinese corpus
annotated with discourse relations. *Journal
of Language Resources and Evaluation*.

Zhou, Y. and N. Xue. 2012. PDTB-style
discourse annotation of Chinese text.
In *Proceedings of the 50th Annual Meeting
of the ACL*, pages 69–77, Jeju Island.

Zhou, Z.-M., M. Lan, Y. Xu, Z.-Y. Niu, J. Su,
and C. L. Tan. 2010. Predicting discourse
connectives for implicit discourse relation
recognition. In *Proceedings of the 23rd
International Conference on Computational
Linguistics (COLING)*, pages 1,507–1,514,
Beijing.