The Penn Discourse Treebank 2.0 Annotation Manual

The PDTB Research Group

December 17, 2007

Contributors:

Rashmi Prasad, Eleni Miltsakaki, Nikhil Dinesh, Alan Lee, Aravind Joshi
Department of Computer and Information Science
and
Institute for Research in Cognitive Science,
University of Pennnsylvania
{rjprasad,elenimi,nikhild,aleewk,joshi}@seas.upenn.edu

and

Livio Robaldo
Department of Informatics
University of Torino
livio.robaldo@di.unito.it

and

Bonnie Webber
School of Informatics,
University of Edinburgh
bonnie@inf.ed.ac.uk

Contents

1 Introduction **1**

1.1 Background and overview . . . 1
1.2 Source corpus and annotation styles . . . 2
1.3 Summary of annotations . . . 3
1.4 Differences between PDTB-1.0. and PDTB-2.0 . . . 4
1.5 A note on multiple connectives . . . 5
1.6 Recommendations for training and testing experiments with PDTB-2.0. . . . 5
1.7 Notation conventions . . . 6

2 `Explicit` connectives and their arguments **8**

2.1 Identifying `Explicit` connectives . . . 8
2.2 Modified connectives . . . 9
2.3 Parallel connectives . . . 10
2.4 Conjoined connectives . . . 10
2.5 Linear order of connectives and arguments . . . 10
2.6 Location of arguments . . . 11
2.7 Types and extent of arguments . . . 12
2.7.1 Simple clauses . . . 12
2.7.2 Non-clausal arguments . . . 13
2.7.2.1 VP coordinations . . . 13
2.7.2.2 Nominalizations . . . 13
2.7.2.3 Anaphoric expressions denoting abstract objects . . . 13
2.7.2.4 Responses to questions . . . 14
2.7.3 Multiple clauses/sentences, and the Minimality Principle . . . 14
2.8 Conventions . . . 14
2.8.1 Clause-internal complements and non-clausal adjuncts . . . 15
2.8.2 Punctuation . . . 16

3 `Implicit` connectives and their arguments **17**

3.1 Introduction . . . 17
3.2 Unannotated implicit relations . . . 18
3.2.1 Implicit relations across paragraphs . . . 18
3.2.2 Intra-sentential relations . . . 18
3.2.3 Implicit relations in addition to explicitly expressed relations . . . 19

3.2.4 Implicit relations between non-adjacent sentences . . . 19
3.3 Extent of arguments . . . 20
3.3.1 Sub-sentential arguments . . . 20
3.3.2 Multiple sentence arguments . . . 20
3.3.3 Arguments involving parentheticals . . . 22
3.4 Non-insertability of `Implicit` connectives . . . 22
3.4.1 `AltLex` (Alternative lexicalization) . . . 22
3.4.2 `EntRel` (Entity-based coherence) . . . 23
3.4.3 `NoRel` (No relation) . . . 25

4 Senses **26**
4.1 Introduction . . . 26
4.2 Hierarchy of senses . . . 26
4.3 Class: "TEMPORAL" . . . 27
4.3.1 Type: "Asynchronous" . . . 28
4.3.2 Type: "Synchronous" . . . 28
4.4 Class: "CONTINGENCY" . . . 28
4.4.1 Type: "Cause" . . . 28
4.4.2 Type: "Pragmatic Cause" . . . 29
4.4.3 Type: "Condition" . . . 29
4.4.4 Type: "Pragmatic Condition" . . . 31
4.5 Class: COMPARISON . . . 32
4.5.1 Type: "Contrast" . . . 32
4.5.2 Type: "Pragmatic Contrast" . . . 33
4.5.3 Type: "Concession" . . . 34
4.6 Class: "EXPANSION" . . . 34
4.6.1 Type: "Instantiation" . . . 34
4.6.2 Type: "Restatement" . . . 35
4.6.3 Type: "Alternative" . . . 36
4.6.4 Type: "Exception" . . . 37
4.6.5 Type: "Conjunction" . . . 37
4.6.6 Type: "List" . . . 37
4.7 Notes on a few connectives . . . 37
4.7.1 Connective: *As if* . . . 38
4.7.2 Connective: *Even if* . . . 38
4.7.3 Connective: *Otherwise* . . . 38

4.7.4 Connectives: *Or* and *when* . . . 39
4.7.5 Connective: *So that* . . . 39

5 Attribution **40**
5.1 Introduction . . . 40
5.2 Source . . . 41
5.3 Type . . . 44
5.3.1 Assertion proposition AOs and belief propositions AOs . . . 44
5.3.2 Fact AOs . . . 45
5.3.3 Eventuality AOs . . . 45
5.4 Scopal polarity . . . 46
5.5 Determinacy . . . 47
5.6 Attribution spans . . . 47

6 Description of PDTB representation format **50**
6.1 Introduction . . . 50
6.2 Directory structure and linking mechanism . . . 50
6.3 File format . . . 51
6.3.1 General outline . . . 52
6.3.2 `Explicit` relation . . . 53
6.3.3 `AltLex` relation . . . 56
6.3.4 `Implicit` relation . . . 58
6.3.5 `EntRel` and `NoRel` . . . 61
6.4 Computation of Gorn addresses . . . 62
6.5 Acknowledgements . . . 64

Appendix A **65**

Appendix B **71**

Appendix C **77**

Appendix D **81**

Appendix E **87**

Appendix F **91**

Appendix G **92**

Appendix H **95**

References **97**

1 Introduction

1.1 Background and overview

An important aspect of discourse understanding and generation involves the recognition and processing of *discourse relations*. Building on some early work on discourse structure in Webber and Joshi (1998), where discourse connectives as treated as discourse-level predicates that take two *abstract objects* such as events, states, and propositions (Asher, 1993) as their arguments, the Penn Discourse Treebank (PDTB) has annotated the *argument structure*, *senses* and *attribution* of discourse connectives and their arguments.[1]

This report documents the annotation guidelines and annotation styles for the second release of the PDTB (PDTB-2.0).[2] The PDTB-2.0. distribution is available through the Linguistic Data Consortium (LDC)[3], and contains the corpus, annotation manuals, relevant publications as well as software to enable some simple and fast processing of the corpus data. PDTB-2.0 contains extensions and revisions of some aspects of the annotation since the first release, primarily with respect to the senses of connectives (Section 4) and the attribution of connectives and their arguments (Section 5).

Discourse connectives in the PDTB include: `Explicit` discourse connectives, which are drawn primarily from well-defined syntactic classes, and `Implicit` discourse connectives, which are inserted between paragraph-internal adjacent sentence pairs not related explicitly by any of the syntactically-defined set of `Explicit` connectives. In the latter case, the reader must attempt to infer a discourse relation between the adjacent sentences, and "annotation" consists of *inserting* a connective expression that *best* conveys the inferred relation. Connectives *inserted* in this way to express inferred relations are called `Implicit` connectives. Multiple discourse relations (Webber *et al.*, 1999) can also be inferred, and are annotated by inserting multiple `Implicit` connectives.

Adjacent sentence-pairs between which annotators found no `Implicit` connective to be appropriate are further distinguished as: (a) `AltLex`, where a discourse relation is inferred, but insertion of an `Implicit` connective leads to *redundancy* in its expression due to the relation being *alternatively lexicalized* by some other expression; (b) `EntRel`, where no discourse relation can be inferred and where the second sentence only serves to provide some further description of an entity in the first sentence (akin to *entity-based coherence* (Knott *et al.*, 2001)); and (c) `NoRel`, where neither a discourse relation nor entity-based coherence can be inferred between the adjacent sentences.

Because there are no generally accepted abstract semantic categories for classifying the arguments to discourse connectives as have been suggested for verbs (e.g., agent, patient, theme, etc.), the two arguments to a discourse connective are simply labelled `Arg2`, for the argument that appears in the clause that is syntactically bound to the connective, and `Arg1`, for the other argument.[4]

[1]The Penn Discourse Treebank Project (`http://www.seas.upenn.edu/~pdtb`) was partially supported by NSF Grant: Research Resources, EIA 02-24417 to the University of Pennsylvania (PI: Aravind Joshi).

[2]In April 2006, a preliminary version of PDTB (PDTB-1.0.) was released in order to get some feedback. This version is no longer available.

[3]`http://www.ldc.upenn.edu/`

[4]All connectives annotated in the PDTB have two and only two arguments. PDTB discourse-level predicate-argument structures are therefore unlike the predicate-argument structures of verbs at the sentence-level (PROPBANK, (Palmer *et al.*, 2005)), where verbs can take any number of arguments. At the same time, however, we note that certain types of constructions could be possibly viewed as structures with more than two arguments, such as "Lists"

Supplements to `Arg1` and `Arg2`, called `Sup1` for material supplementary to `Arg1`, and `Sup2`, for material supplementary to `Arg2`, are annotated to mark material that is relevant but not "mininally necessary" for interpretating the relation.

Annotation of `Explicit` connectives and their arguments consists of selecting the corresponding span of text in the source text files. Supplementary material is annotated in the same way. `Implicit` connectives are annotated by first selecting the first character of `Arg2` as the textual span for the `Implicit` connective, then selecting the text spans for `Arg1` and `Arg2` of the relation, and finally providing a word or phrase to express the relation. In the case of `AltLex`, instead of providing a word/phrase, the text span in `Arg2` expressing the relation is selected and marked. `EntRel` and `NoRel` annotations only involve selection of the first character of `Arg2` as the placeholder for the relation and then selection of the adjacent sentences as `Arg1` and `Arg2`.

Senses of connectives are annotated for `Explicit` connectives, `Implicit` connectives and `Altlex` relations. No senses are provided for `EntRel` and `NoRel` since no discourse relations are inferred for these. Sense labels are drawn from a hierarchical classification - a three-level hierarchy grouping connectives into *classes*, *types* and *subtypes* - and are annotated as features on connectives.

Attribution, which is a relation of "ownership" between individuals and abstract objects, is annotated for `Explicit` connectives, `Implicit` connectives and `Altlex` relations, as well as their arguments. The annotation scheme aims to capture both the source and degrees of factuality of the abstract objects through the annotation of text spans signalling the attribution, and of features recording the *source*, *type*, *scopal polarity*, and *determinacy* of attribution.

The annotation guidelines described in this document draw and expand on earlier reports presented in annotation tutorials and papers, notably Miltsakaki *et al.* (2004a,b); Prasad *et al.* (2004); Dinesh *et al.* (2005); Prasad *et al.* (2005); Webber *et al.* (2005); Miltsakaki *et al.* (2005); Prasad *et al.* (2006, 2007). The rest of this section discusses the source corpus and annotation style of PDTB-2.0, and presents an overview of the annotation contained in the corpus, including an overview of the extensions from PDTB-1.0. Section 2 presents the annotation guidelines for the argument structure of `Explicit` connectives. Annotation guidelines for implicit relations and their arguments are presented in Section 3. Section 4 presents the guidelines for sense annotation. Section 5 describes the guidelines for attribution annotation. File structures and representation formats of the corpus are described in Section 6. Finally, Appendices A-H provide distributions of some aspects of the annotations.

1.2 Source corpus and annotation styles

The PDTB annotations are done on the same Wall Street Journal (WSJ) corpus on which the Penn Treebank (PTB) II corpus (Marcus *et al.*, 1993) was built. Annotation of connectives and their arguments consists of recording the text spans that anchor them in the WSJ RAW files, but the final annotation representation follows the "stand-off" annotation technique, such that the text spans are

and logical "if-then-else" constructions. In the PDTB, "Lists" with more than two elements are composed of multiple binary structures; that is, every new successive element of the List forms one of the arguments of a "new" relation, with all prior elements together forming the second argument. While we would have preferred to structure Lists differently, the reason for annotating them as they are is that their definition and reliable identification proved to be problematic. As for "if-then-else" constructions, we did not find any clear cases of these constructs in the corpus.

represented in terms of their character offsets in the WSJ RAW text files. To enable some simple and fast processing of the data, the "raw text" associated with the text spans is also given as a feature on the primary stand-off annotation.

All text spans annotated in the PDTB (i.e., the spans for connectives and their arguments, the spans for `AltLex` relations, and the spans for attribution) are also linked to the "parsed" PTB files in a similar stand-off style, with the reference to the PTB structural description being represented as a set of tree node *Gorn* address.[5] Other aspects of the annotation, including sense and attribution, are represented as features. A complete description of the representation format of the PDTB annotations is provided in Section 6.

Because of the stand-off annotation style of the PDTB, the corpus can be effectively used only in conjunction with the primary source data, i.e., the WSJ RAW and PTB parsed files, which must be obtained independently from the LDC PTB-II distribution.[6]

Since the PDTB provides links to the PTB parsed files, only those 2304 RAW files that have corresponding parsed files were chosen for PDTB annotations. Of these 2304 files, only 2159 files are contained in the PDTB distribution: 8 of the 2304 parsed files could not be converted to stand-off, and the remaining texts did not have any occurrences of discourse relations that could be annotated with the PDTB scheme.[7]

1.3 Summary of annotations

There are a total of 40600 tokens annotated in PDTB-2.0, annotated variously as (see Table 1 for WSJ section-wise distributions):

1. `Explicit` connectives, of which there are 18459 tokens and 100 distinct types. Modified connectives such as *only because*, *just when*, etc. are treated as belonging to the same type as that of their head (i.e., *because*, *when*, etc.). 111 distinct senses are recorded for `Explicit` connectives, with multiple senses provided for a connective together treated as a distinct sense. Appendix A gives the distribution of all the distinct types of `Explicit` connectives and the set of senses annotated for each type. Appendix B gives the distribution of all the distinct senses annotated for `Explicit` connectives and the set of connectives associated with each type. The distribution of the full forms of `Explicit` connectives, including modified and non-modified forms is given in Appendix C. While attribution is annotated for all relations, explicit attributions appear in 6452 (34%) `Explicit` connective tokens. Appendix G gives the distribution of distinct attribution feature values annotated for all `Explicit` connectives, with and without explicit attributions.

2. `Implicit` connectives, of which there are 16053 tokens. Multiple connectives are provided for 171 tokens, and when each element of a multiple connective is counted separately, the total number of tokens is 16224, with 102 distinct types. (See further discussion of multiple

[5] The links to the PTB parsed texts were generated programmatically. We have used these links in our experiments, but all have not been examined by a human.

[6] `http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC95T7`

[7] The 8 parsed files that could not be converted to stand-off are: wsj_0203, wsj_0285, wsj_0455, wsj_0749, wsj_0998, wsj_1625, wsj_2170, wsj_2312.

SECTION	Explicit	Implicit	AltLex	EntRel	NoRel	TOTAL
Sec. 00	712	592	19	218	30	1571
Sec 01	750	591	20	271	3	1635
Sec. 02	713	708	28	200	8	1657
Sec. 03	529	446	13	166	2	1156
Sec. 04	822	747	27	238	5	1839
Sec. 05	816	780	29	148	15	1788
Sec. 06	653	571	14	143	13	1394
Sec. 07	804	676	24	302	5	1811
Sec. 08	176	161	2	56	8	403
Sec. 09	786	720	13	193	31	1743
Sec. 10	720	613	5	185	15	1538
Sec. 11	780	839	35	208	16	1878
Sec. 12	800	726	35	180	16	1757
Sec. 13	941	863	32	250	8	2094
Sec. 14	734	731	31	244	13	1753
Sec. 15	868	703	40	201	5	1817
Sec. 16	1092	993	61	243	5	2394
Sec. 17	614	487	22	201	6	1330
Sec. 18	898	722	32	265	7	1924
Sec. 19	647	519	34	250	7	1457
Sec. 20	724	627	20	257	14	1642
Sec. 21	605	524	27	203	5	1364
Sec. 22	680	522	19	215	8	1444
Sec. 23	923	769	30	217	4	1943
Sec. 24	672	423	12	156	5	1268
ALL	18459	16053	624	5210	254	40600

Table 1: Section-wise distribution of PDTB-2.0. annotations

connectives in Section 1.5.) 84 distinct senses are recorded for `Implicit` connectives, with multiple senses provided for a connective treated together as a distinct sense. Appendix D gives the distribution of all the distinct types of `Implicit` connectives and the set of senses annotated for each type. Appendix E gives the distribution of all the distinct senses annotated with `Implicit` connectives and the set of connectives associated with each sense. While attribution is annotated for all relations, explicit attributions appear in 5711 (35%) `Implicit` connective tokens. Appendix G gives the distribution of distinct attribution feature values annotated for all `Implicit` connectives and `AltLex` relations (see below), with and without explicit attributions.

3. `AltLex`, of which there are 624 tokens, with 28 distinct senses. Appendix F gives the distribution of the senses annotated for `AltLex`. Explicit attributions appear in 187 (29%) tokens.

4. `EntRel`, of which there are 5210 tokens. (Senses and attributions are not marked for `EntRel`.)

5. `NoRel`, of which there are 254 tokens. (Senses and attributions are not marked for `NoRel`.)

1.4 Differences between PDTB-1.0. and PDTB-2.0

Differences between PDTB-1.0 and PDTB-2.0 include:

1. Annotations of implicit relations across the entire corpus (provided in PDTB-1.0 for only 3 WSJ sections).
2. Revision of the annotation scheme for senses (PDTB-1.0 used a simpler 7-way broad classification).
3. Sense annotations of all `Explicit` connectives, `Implicit` connectives, and `AltLex` relations (provided in PDTB-1.0 for only implicit relations).
4. Revision of the annotation scheme for attribution (more detailed than the scheme used in PDTB-1.0).

1.5 A note on multiple connectives

Where multiple `Explicit` connectives occur in the same clause (either two discourse adverbials or a conjunction and a discourse adverbial, such as "so for instance", "but then", "and furthermore", "previously for example", etc.), each connective is annotated separately, taking its own two arguments. However, this ignores the real possibility that the connectives are **not** independent – that one connective depends on the other, as in "`Arg1` because for instance `Arg2`", where `Arg2` exemplifies not `Arg1` but a reason for `Arg1`, or in "`Arg1`. Previously for example `Arg2`", where `Arg2` exemplifies not `Arg1` but an event or situation previous to `Arg1` (Webber *et al.*, 2003).

As the PDTB annotation does not explicitly distinguish dependent from independent connectives when they appear together in a clause (either as differences in argument span or as differences in sense annotation), the distinction between the two has also been ignored in our tabulations (cf. Appendix A, B, D, and E). However, any work on *inducing* recognizers for connectives, their arguments and their senses from contextual features should recognize that ignoring the difference between independent and dependent connectives may lead to features being less than predictive, leading in turn to less accurate classifiers.

Multiple `Implicit` connectives have also been annotated in the PDTB. Unlike the multiple `Explicit` connectives, these are not annotated as separate tokens, i.e., both are together associated with the same set of arguments, although sense annotations are provided for each of the connectives. However, here too, no distinction is made between dependent and independent connectives, either in terms of their argument structure, or their sense. As such, the caveat noted for multiple `Explicit` connectives above must be heeded for multiple `Implicit` connectives as well.

In the distributions provided in Appendix A, B, D, and E, we have counted both multiple `Explicit` and multiple `Implicit` connectives separately, in order to be consistent.

1.6 Recommendations for training and testing experiments with PDTB-2.0.

Here we make some recommendations to researchers intending to use the Penn Discourse TreeBank (PDTB-2.0) for corpus-based machine learning (eg, for inducing recognisers for `Implicit` connectives,

sense taggers for Explicit connectives, discourse parsers, etc.).

Statistical experimental work on parsing using the Penn TreeBank (PTB) has been based on using Sections 2 to 21 for training, Section 22 for development and Section 23 for testing. This choice of sections for training, development and testing has historical roots in early parsing experiments carried out by David Magerman. Subsequent researchers have continued this selection of sections, thus making it possible to compare experimental results. Sections 0, 1 and 24 have not been used except in some recent work, where Section 24 has been used as an additional development section.

For experiments with the PDTB, we recommend that Section 2 to 21 be used for training, Section 22 for development and Section 23 for testing. Further, Sections 0, 1, and 24 are recommended as additional development sections, if desired.

Our recommendation is based on the following considerations. First, since there is already a tradition of using certain sections for training, development and testing for parsing, we do not at this time see any harm in continuing this tradition. Second, the section-wise counts (see Table 1) of the PDTB annotations, especially for Explicit and Implicit connectives, do not show any radically skewed distributions (except for Section 08, but this is because there are only 21 files in this Section).

We also suggest that anyone attempting specialized experiments on the corpus, such as for attribution, senses, etc., should carry out a detailed distributional analysis of the corpus for the phenomenon in question before simply adopting the standard practice. They should then publish and share that analysis with others, so that appropriate "best practice" becomes shared.

1.7 Notation conventions

In all examples, the parenthesized 4-digit number after an example gives the WSJ RAW file number containing the example. Annotated Explicit connectives are underlined, and annotated Implicit connectives are shown in small caps. For clarity, Implicit connectives are further indicated by the marker, "Implicit =". For the arguments of connectives, the text whose interpretation is the basis for Arg1 appears in italics, while that of Arg2 appears in bold. For example, in (1), the subordinating conjunction *because* is an Explicit connective that establishes a causal relation between *the campaign board refusing to pay Mr. Dinkins* (Arg1) and *Mr. Dinkin's campaign records being incomplete* (Arg2). In Example (2), the Implicit connective *so* has been inserted to express the inferred consequence relation between the second and third sentences, i.e., between *Motorola no longer delivering junk mail* (Arg1) and *the mail going into the trash* (Arg2).

(1) *The city's Campaign Finance Board has refused to pay Mr. Dinkins $95,142 in matching funds* because **his campaign records are incomplete**. (0041)

(2) Motorola is fighting back against junk mail. *So much of the stuff poured into its Austin, Texas, offices that its mail rooms there simply stopped delivering it.* Implicit = SO **Now, thousands of mailers, catalogs and sales pitches go straight into the trash**. (0989)

AltLex, EntRel, and NoRel annotations are also indicated by underlining, i.e., as "AltLex" (Example 3), "EntRel" (Example 4), and "NoRel" (Example 5). The elsewhere lexicalizing expression for AltLex is shown in square brackets.

(3) *After trading at an average discount of more than 20% in late 1987 and part of last year, country funds currently trade at an average premium of 6%.* AltLex [**The reason:**] **Share prices of many of these funds this year have climbed much more sharply than the foreign stocks they hold**. (0034)

(4) *Pierre Vinken, 61 years old, will join the board as a nonexecutive director Nov. 29.* EntRel **Mr. Vinken is chairman of Elsevier N.V., the Dutch publishing group**. (0001)

(5) *Mr. Rapanelli met in August with U.S. Assistant Treasury Secretary David Mulford.* NoRel **Argentine negotiator Carlos Carballo was in Washington and New York this week to meet with banks**. (0021)

Supplementary annotations are shown in subscripted parentheses, as seen for `Sup1` in Example (6).

(6) ($_{Sup1}$ Workers described "clouds of blue dust") *that hung over parts of the factory*, even though **exhaust fans ventilated the area**. (0003)

In Section 4, we show the annotation of senses for connectives, where they appear at the end of the example in parentheses and reflect the hierarchical classification of the annotated sense, as shown in Example (7).

(7) *Use of dispersants was approved* when **a test on the third day showed some positive results**, officials said. (CONTINGENCY:Cause:reason) (1347)

2 `Explicit` connectives and their arguments

2.1 Identifying `Explicit` connectives

`Explicit` connectives in the PDTB are drawn from the following grammatical classes:

- *Subordinating conjunctions* (e.g., *because*, *when*, *since*, *although*):

  (8) Since **McDonald's menu prices rose this year**, *the actual decline may have been more.* (1280)

  (9) *The federal government suspended sales of U.S. savings bonds* because **Congress hasn't lifted the ceiling on government debt**. (0008)

- *Coordinating conjunctions* (e.g., *and*, *or*, *nor*):[8]

  (10) *The House has voted to raise the ceiling to $3.1 trillion*, but **the Senate isn't expected to act until next week at the earliest**. (0008)

  (11) The report offered new evidence that the nation's export growth, though still continuing, may be slowing. *Only 19% of the purchasing managers reported better export orders in October, down from 27% in September.* And **8% said export orders were down last month, compared with 6% the month before**. (0036)

- *(ADVP and PP) adverbials* (e.g., *however*, *otherwise*, *then*, *as a result*, *for example*).[9]

  (12) *Working Woman, with circulation near one million, and Working Mother, with 625,000 circulation, are legitimate magazine success stories.* **The magazine Success**, however, **was for years lackluster and unfocused**. (1903)

  (13) *In the past, the socialist policies of the government strictly limited the size of new steel mills, petrochemical plants, car factories and other industrial concerns to conserve resources and restrict the profits businessmen could make.* As a result, **industry operated out of small, expensive, highly inefficient industrial units**. (0629)

Adverbials that do not denote relations between two abstract objects (AOs) have not been annotated as discourse connectives. For example, adverbials called "cue phrases" or "discourse markers" such as *well*, *anyway*, *now*, etc. (Hirschberg and Litman, 1987), have not been annotated since they serve to signal the organizational or focus structure of the discourse, rather than relate AOs. And clausal

[8] Only coordinating conjunctions between clauses have been annotated. Coordinating conjunctions appearing in VP coordinations have not been annotated, such as the conjunction *and* in (i) below:

(i) More common chrysotile fibers are curly *and* are more easily rejected by the body, Dr. Mossman explained. (0003)

[9] The adverbials *in fact* and *indeed* were annotated as a discourse connective, although we now think (on theoretical grounds) that it is probably not one (Forbes-Riley *et al.*, 2006). We will be examining the annotation of *in fact* to see whether there is empirical evidence to back up this theoretically motivated decision.

adverbials such as *strangely*, *probably*, *frankly*, *in all likelihood* etc. are also not annotated as discourse connectives since they take a single AO as argument, rather than two (Forbes-Riley *et al.*, 2006).

Not all tokens of words and phrases that can serve as `Explicit` connectives (see Appendix A and C) actually do so: Some tokens serve other functions, such as to relate non-AO entities (e.g., the use of *and* to conjoin noun phrases in Example (14), and the use of *for example* to modify a noun phrase in Example (15)), to relativize extracted adjuncts (e.g., the use of *when* to relativize the time NP in Example (16)), and so on. Such expressions are not annotated as discourse connectives.

(14) Dr. Talcott led a team of researchers from the National Cancer Institute *and* the medical schools of Harvard University *and* Boston University. (0003)

(15) These mainly involved such areas as materials – advanced soldering machines, *for example* – and medical developments derived from experimentation in space, such as artificial blood vessels. (0405)

(16) Equitable of Iowa Cos., Des Moines, had been seeking a buyer for the 36-store Younkers chain since June, *when* it announced its intention to free up capital to expand its insurance business. (0156)

2.2 Modified connectives

Many connectives can occur with adverbs such as *only*, *even*, *at least*, and so on. We refer to such tokens as *modified connectives* (with the connective as *head* and the adverb as *modifier*). Some examples are given in Examples (17-19), with the adverb shown in parentheses for clarity. Rather than distinguishing such occurrences as a separate type, they are treated as the same type as that of the head - the bare form.[10] Appendix C lists all modified connectives found and annotated in the PDTB corpus for each connective type.

(17) *That power can sometimes be abused*, (particularly) since **jurists in smaller jurisdictions operate without many of the restraints that serve as corrective measures in urban areas**. (0267)

(18) *You can do all this* (even) if **you're not a reporter or a researcher or a scholar or a member of Congress**. (0108)

(19) *We're seeing it* (partly) because **older vintages are growing more scarce**. (0071)

While we have annotated modified connectives such as described above, certain types of post-modified connectives have not been annotated, in particular those post-modified by prepositions, for example *because (of)...*, *as a result (of)...*, *instead (of)...*, and *rather (than)....* While in many cases such expressions relate noun phrases lacking an AO interpretation, (Example 20), there are also a few cases such as Example 21 where they do relate AOs. However, these few tokens have not been annotated.

[10] In the corpus, the head of a modified connective is given as a feature. The value of that feature for unmodified connectives is the connective itself. See the description of PDTB representation formats (Section 6).

(20) The products already available are cross-connect systems, used *instead of* mazes of wiring to interconnect other telecommunications equipment. (1064)

(21) *instead of* featuring a major East Coast team against a West Coast team, it pitted the Los Angeles Dodgers against the losing Oakland A's. (0443)

2.3 Parallel connectives

In addition to modified forms of connectives, we have also annotated a small set of "parallel" connectives, that is, pairs of connectives where one part presupposes the presence of the other, and where both together take the same two arguments (Examples (22-24)). Such connectives are listed as distinct types and are annotated discontinuously.[11]

(22) On the one hand, Mr. Front says, *it would be misguided to sell into "a classic panic."* On the other hand, **it's not necessarily a good time to jump in and buy**. (2415)

(23) If *the answers to these questions are affirmative*, then **institutional investors are likely to be favorably disposed toward a specific poison pill**. (0275)

(24) Either *sign new long-term commitments to buy future episodes* or **risk losing "Cosby" to a competitor**. (0060)

2.4 Conjoined connectives

Conjoined connectives like *when and if* and *if and when* are treated as complex connectives and listed as distinct types. Examples are shown in (25-26).

(25) When and if **the trust runs out of cash** – which seems increasingly likely – *it will need to convert its Manville stock to cash.* (1328)

(26) Hoylake dropped its initial $13.35 billion ($20.71 billion) takeover bid after it received the extension, but said *it would launch a new bid* if and when **the proposed sale of Farmers to Axa receives regulatory approval**. (2403)

2.5 Linear order of connectives and arguments

Connectives and their arguments can appear in any relative order. For the subordinating conjunctions, since the subordinate clause is bound to the connective, `Arg2` corresponds to the subordinate clause, and hence the linear order of the arguments can be `Arg1-Arg2` (Ex. 27), `Arg2-Arg1` (Ex. 28), or `Arg2` may appear between discontinuous parts of `Arg1` (Ex. 29), depending on the relative position of the subordinate clause with respect to its matrix clause.

(27) *The federal government suspended sales of U.S. savings bonds* because **Congress hasn't lifted the ceiling on government debt**. (0008)

[11] In Appendices A, B and C, parallel connectives are shown with two dots between the two parts of the pair (e.g., *on the one hand..on the other hand*, *if..then*, *either..or*).

(28) Because **it operates on a fiscal year**, *Bear Stearns's yearly filings are available much earlier than those of other firms.* (1948)

(29) *Most oil companies*, when **they set exploration and production budgets for this year**, *forecast revenue of $15 for each barrel of crude produced.* (0725)

The order of the arguments for adverbials and coordinating conjunctions is typically `Arg1-Arg2` since `Arg1` usually appears in the prior discourse. But as Example (30) shows, `Arg1` of a discourse adverbial can also appear within `Arg2`, which is then annotated as two discontinuous spans.

(30) As an indicator of the tight grain supply situation in the U.S., market analysts said **that late Tuesday the Chinese government**, *which often buys U.S. grains in quantity*, **turned** instead **to Britain to buy 500,000 metric tons of wheat**. (0155)

The position of connectives in the `Arg2` clause they modify is restricted to initial position for subordinating and coordinating conjunctions, but adverbials may occur medially or finally in `Arg2`, as shown below:

(31) *Despite the economic slowdown, there are few clear signs that growth is coming to a halt.* As a result, **Fed officials may be divided over whether to ease credit**. (0072)

(32) *The chief culprits*, he says, *are big companies and business groups that buy huge amounts of land "not for their corporate use, but for resale at huge profit."* ... **The Ministry of Finance**, as a result, **has proposed a series of measures that would restrict business investment in real estate** ... (0761)

(33) *Polyvinyl chloride capacity "has overtaken demand* **and we are experiencing reduced profit margins** as a result", ... (2083)

2.6 Location of arguments

There is no restriction on how far an argument can be from its corresponding connective. So arguments can be found in the same sentence as the connective (Examples 34-36), in the sentence immediately preceding that of the connective (Examples 37-39), or in some non-adjacent sentence (Example 40).

(34) *The federal government suspended sales of U.S. savings bonds* because **Congress hasn't lifted the ceiling on government debt**. (0008)

(35) *Most balloonists seldom go higher than 2,000 feet* and **most average a leisurely 5-10 miles an hour**. (0239)

(36) *In an invention that drives Verdi purists bananas, Violetta lies dying in bed during the prelude*, rising deliriously **when** then **she remembers the great parties she used to throw**. (1154)

(37) *Why do local real-estate markets overreact to regional economic cycles*? Because **real-estate purchases and leases are such major long-term commitments that most companies**

and individuals make these decisions only when confident of future economic stability and growth. (2444)

(38) *Metropolitan Houston's population has held steady over the past six years.* And **personal income, after slumping in the mid-1980s, has returned to its 1982 level in real dollar terms**. (2444)

(39) *Such problems will require considerable skill to resolve.* However, **neither Mr. Baum nor Mr. Harper has much international experience**. (0109)

(40) Mr. Robinson of Delta & Pine, the seed producer in Scott, Miss., said *Plant Genetic's success in creating genetically engineered male steriles doesn't automatically mean it would be simple to create hybrids in all crops.* (`Sup1` That's because pollination, while easy in corn because the carrier is wind, is more complex and involves insects as carriers in crops such as cotton). "It's one thing to say you can sterilize, and another to then successfully pollinate the plant," he said. Nevertheless, he said, **he is negotiating with Plant Genetic to acquire the technology to try breeding hybrid cotton**. (0209)

2.7 Types and extent of arguments

2.7.1 Simple clauses

With a few exceptions to be discussed below (Section 2.7.2), the simplest syntactic realization of an abstract object as a connective's argument is taken to be a clause, tensed or non-tensed. Further, the clause can be a matrix clause, a complement clause, or a subordinate clause. Some examples of single clausal realizations are shown in Examples (41-46). For clause types such as non-finite clauses and relative clauses, the argument selection assumes the presence of implicit subjects and traces of extracted complements available in the syntactic structure of the clause in the PTB, so that a complete description of the argument can be derived from the links to the PTB.

(41) A Chemical spokeswoman said *the second-quarter charge was "not material"* **and that no personnel changes were made** as a result. (0304)

(42) In Washington, House aides said Mr. Phelan told congressmen that the collar, *which banned program trades through the Big Board's computer* when **the Dow Jones Industrial Average moved 50 points**, didn't work well. (0088)

(43) *Knowing a tasty – and free – meal* when **they eat one**, the executives gave the chefs a standing ovation. (0010)

(44) Alan Smith, president of Marks & Spencer North America and Far East, says that Brooks Brothers' focus is to boost sales *by broadening its merchandise assortment* while **keeping its "traditional emphasis."** (0530)

(45) Radio Shack says it has a policy *against selling products* if **a salesperson suspects they will be used illegally**. (1058)

(46) "We have been a great market *for inventing risks* **which other people** then **take, copy and cut rates**." (1302)

2.7.2 Non-clausal arguments

In some exceptional cases, non-clausal elements are treated as realizations of abstract objects.

2.7.2.1 VP coordinations While the conjunction in a coordinated verb phase is not annotated as a distinct discourse connective (cf. footnote 8), one or more verb phrases within the coordinated structure can be annotated as the argument of another connective. However, the subject of the VP coordinates is included in the argument selection only for the first VP coordinate (`Arg1` of *then* in Example 47). Subjects for non-initial coordinates are not included in the selection (`Arg2` of *then* in Example 47 and `Arg1` of *because* in Example 48), and will have to be retrieved via independent heuristics to arrive at the complete interpretation of the argument.

(47) *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). (0091)

(48) She became an abortionist accidentally, *and continued* because **it enabled her to buy jam, cocoa and other war-rationed goodies**. (0039)

2.7.2.2 Nominalizations Nominalizations are annotated as arguments of connectives in two strictly restricted contexts. The first context is when they allow for an *existential* interpretation, as in Example (49), where the `Arg1` selection can be interpreted existentially as *that there will be major new liberalizations*:

(49) Economic analysts call his trail-blazing liberalization of the Indian economy incomplete, and many are hoping *for major new liberalizations* if **he is returned firmly to power**. (2041)

The second context is when they involve a clearly observable case of a *derived nominalization*, as in Example (50), where the `Arg1` selection can be assumed to be transformationally derived from *such laws to be resurrected*:

(50) But in 1976, the court permitted *resurrection of such laws*, if **they meet certain procedural requirements**. (0426)

2.7.2.3 Anaphoric expressions denoting abstract objects An anaphoric expression like *this* or *that* or *so* that refers to an abstract object can be annotated as `Arg1` of a connective. With such annotation we assume that an anaphora resolution mechanism will yield the interpretation of the argument.[12]

(51) "It's important to share the risk *and even more so* when **the market has already peaked**." (0782)

[12] In some cases, the (AO) antecedent of the anaphor has been marked as supplementary material, but this has not been done consistently.

(52) Investors who bought stock with borrowed money – that is, "on margin" – may be more worried than most following Friday's market drop. *That's* <u>because</u> **their brokers can require them to sell some shares or put up more cash to enhance the collateral backing their loans**. (2393)

(53) Evaluations suggest that good ones are – *especially so* <u>if</u> **the effects on participants are counted**. (2412)

2.7.2.4 Responses to questions In some contexts such as question-answer sequences, where the response to a question only includes response particles like *yes* and *no*, the response particles are themselves annotated as arguments, with the preceding question annotated as supplementary material to indicate the question-answer relation.

(54) Underclass youth are a special concern. (`Sup1` Are such expenditures worthwhile, then)? *Yes*, <u>if</u> **targeted**. (2412)

(55) (`Sup1` Is he a victim of Gramm-Rudman cuts)? *No*, <u>but</u> **he's endangered all the same**: His new sitcom on ABC needs a following to stay on the air. (0528)

2.7.3 Multiple clauses/sentences, and the Minimality Principle

In addition to single clauses, abstract object arguments of connectives can also be realized as multiple clauses and multiple sentences. Example (56) shows multiple sentences selected for the `Arg1` argument of *still*. Multiple clause and multiple sentence arguments can also be annotated discontinuously if they so appear in the text.

(56) *Here in this new center for Japanese assembly plants just across the border from San Diego, turnover is dizzying, infrastructure shoddy, bureaucracy intense. Even after-hours drag; "karaoke" bars, where Japanese revelers sing over recorded music, are prohibited by Mexico's powerful musicians union.* <u>Still</u>, **20 Japanese companies, including giants such as Sanyo Industries Corp., Matsushita Electronics Components Corp. and Sony Corp. have set up shop in the state of Northern Baja California**. (0300)

There are no restrictions on how many or what types of clauses can be included in these complex selections, except for the *Minimality Principle*, according to which only as many clauses and/or sentences should be included in an argument selection as are *minimally required* and *sufficient* for the interpretation of the relation. Any other span of text that is perceived to be relevant (but not necessary) in some way to the interpretation of arguments is annotated as *supplementary information*, labelled `Sup1` and `Sup2`, for `Arg1` and `Arg2` respectively.

2.8 Conventions

This section describes certain conventions that we have followed in the annotation. For such cases, we do not make any claims about whether and how they contribute to the interpretation of the relations. They were mostly adopted for convenience of annotation.

2.8.1 Clause-internal complements and non-clausal adjuncts

For all clauses that are selected as arguments of connectives, all complements of the main clausal predicate and all non-clausal adjuncts (e.g., *a speciality chemicals concern* in `Arg2` of Example 57), adverbs (e.g., *for example* in `Arg1` of Example 58), complementizers (e.g., *that* in `Arg1` and `Arg2` of Example 59), conjunctions (e.g., *But* in `Arg1` of Example 60), and relative pronouns (e.g., *whom* in `Arg1` of Example 61) modifying the clause are obligatorily included in the argument (except for the connective that is itself being annotated), even if these elements are not necessary for the *minimal* interpretation of the relation (see Section 2.7.3).

(57) Although **Georgia Gulf hasn't been eager to negotiate with Mr. Simmons and NL, a specialty chemicals concern**, *the group apparently believes the company's management is interested in some kind of transaction.* (0080)

(58) players must abide by strict rules of conduct even in their personal lives – *players for the Tokyo Giants, for example, must always wear ties* when **on the road**. (0037)

(59) There seems to be a presumption in some sectors of (Mexico's) government *that there is a lot of Japanese money waiting behind the gate*, and **that by slightly opening the gate, that money will enter Mexico**. (0300)

(60) *But the Reagan administration thought otherwise*, and **so may the Bush administration**. (0601)

(61) That impressed Robert B. Pamplin, Georgia-Pacific's chief executive at the time, *whom Mr. Hahn had met* while **fundraising for the institute**. (0100)

Inclusion of non-clausal elements is obligatory even when it warrants discontinuous annotation (Examples 62-65).

(62) They found students in an advanced class a year earlier who said she gave them similar help, *although* because **the case wasn't tried in court**, *this evidence was never presented publicly.* (0044)

(63) He says *that* when **Dan Dorfman, a financial columnist with USA Today, hasn't returned his phone calls**, *he leaves messages with Mr. Dorfman's office saying that he has an important story on Donald Trump, Meshulam Riklis or Marvin Davis.* (1376)

(64) Under two new features, participants will be able to transfer money from the new funds to other investment funds *or*, if **their jobs are terminated**, *receive cash from the funds.* (0204)

(65) *Last week*, when **her appeal was argued before the Missouri Court of Appeals**, *her lawyer also relied on the preamble.* (1423)

Non-clausal attributing phrases are also included obligatorily in the clausal argument they modify, such as "according to..." phrases in the `Arg1` of both the following examples:

(66) *No foreign companies bid on the Hiroshima project, according to the bureau.* But **the Japanese practice of deep discounting often is cited by Americans as a classic barrier to entry in Japan's market**. (0501)

(67) *Even so, according to Mr. Salmore, the ad was "devastating"* because **it raised questions about Mr. Courter's credibility**. (0041)

Verbs of attribution along with their subject are in general excluded from an argument when the attribution does not itself play a role in the interpretation of the relation (see Section 5). So the constraint *against* excluding non-clausal attribution phrases stands as an exception to the general guideline for attribution annotation. However, since attribution annotation also includes annotation of the *attribution span*, it would be straightforward to strip away such attribution spans whenever they appear in their argument span.

2.8.2 Punctuation

For practical reasons in the annotation process, all punctuation at the boundaries of connective and argument selections was excluded. However, in the annotation links to the PTB parsed files, some heuristics are used to extend the annotation spans to include certain boundary punctuation. So while the text annotation does not include punctuation occurring at the edges of arguments, punctuation can, in some cases, be obtained from the linked annotation. Punctuation heuristics and extensions are described in detail in Section 6.

3 Implicit connectives and their arguments

3.1 Introduction

The goal of annotating Implicit connectives in the PDTB is to capture relations between abstract objects that are not realized explicitly in the text (by one of a set of the lexically-defined Explicit connectives - see Section 2.1) and are left to be inferred by the reader. In Example (68), a causal relation is inferred between *raising cash positions to record levels* and *high cash positions helping to buffer a fund*, even though no Explicit connective appears in the text to express this relation. Similarly, in Example (69), a consequence relation is inferred between *the increase in the number of rooms* and *the increase in the number of jobs*, though no Explicit connective expresses this relation.

(68) Several leveraged funds don't want to cut the amount they borrow because it would slash the income they pay shareholders, fund officials said. But a few funds have taken other defensive steps. *Some have raised their cash positions to record levels.* Implicit = BECAUSE **High cash positions help buffer a fund when the market falls.** (0983)

(69) *The projects already under construction will increase Las Vegas's supply of hotel rooms by 11,795, or nearly 20%, to 75,500.* Implicit = SO **By a rule of thumb of 1.5 new jobs for each new hotel room, Clark County will have nearly 18,000 new jobs**. (0994)

In the PDTB, such inferred relations are annotated between adjacent sentences within the same paragraph, and are marked as Implicit connectives, by the insertion of a connective expression that best expresses the inferred relation. So in Examples (68) and (69), the Implicit connectives *because* and *so* are inserted to capture the perceived causal and consequence relations respectively.

Multiple discourse relations between adjacent sentences may also be inferred, and have been annotated as multiple Implicit connectives. In Example (70), two Implicit connectives, *when* and *for example*, are inserted to express how Arg2 presents one instance of the circumstances under which *Mr. Morishita comes across as an outspoken man of the world.* Similarly, in Example (71), the two Implicit connectives *because* and *for example* are provided to express how Arg2 presents one instance of the reasons for the claim that *the third principal did have garden experience.*

(70) *The small, wiry Mr. Morishita comes across as an outspoken man of the world.* Implicit = WHEN FOR EXAMPLE **Stretching his arms in his silky white shirt and squeaking his black shoes he lectures a visitor about the way to sell American real estate and boasts about his friendship with Margaret Thatcher's son.** (0800)

(71) *The third principal in the S. Gardens adventure did have garden experience.* Implicit = BECAUSE FOR EXAMPLE **The firm of Bruce Kelly/David Varnell Landscape Architects had created Central Park's Strawberry Fields and Shakespeare Garden**. (0984)

The decision to *lexically encode* inferred relations in this way was made with the aim of achieving high reliability among annotators while avoiding the difficult task of training them to reason about pre-defined abstract relations. The annotation of inferred relations was thus done intuitively, and involved reading adjacent sentences (and in some cases, the preceding text as well - see Section 3.3.2), making a decision about whether or not a relation could be inferred between them, and providing an

appropriate `Implicit` connective to express the inferred relation, if any. Three distinct pre-defined labels, `AltLex`, `EntRel` and `NoRel` (Section 3.4), were used for cases where an `Implicit` connective could *not* be provided: `AltLex` for cases where the insertion of an `Implicit` connective to express an inferred relation led to a *redundancy* in the expression of the relation; `EntRel` for cases where only an *entity-based coherence* relation could be perceived between the sentences; and `NoRel` for cases where no discourse relation or entity-based coherence relation could be perceived between the sentences.

`Implicit` connectives are annotated between all successive pairs of sentences within paragraphs (see Section 3.2), but they are also annotated intra-sententially between complete clauses delimited by semi-colon (";") or colon (":").

The complete distribution of the types of `Implicit` connectives in the PDTB, along with their associated senses, are given in Appendix D and E. Appendix F gives the distribution of the senses annotated for `AltLex` relations.

3.2 Unannotated implicit relations

Because of time and resource constraints, there are certain circumstances in which implicit relations have not been annotated where they otherwise could have been.

3.2.1 Implicit relations across paragraphs

While an implicit discourse relation can hold between the final sentence of one paragraph and the initial sentence of the next, implicit relations have not been annotated between adjacent sentences separated by a paragraph boundary.[13] For example, in (292) a causal relation can be inferred between the last sentence of the first paragraph and the first sentence of the second paragraph, in that the latter provides one reason why the 1% charge is in fact the best bargain available.

(72) The Sept. 25 "Tracking Travel" column advises readers to "Charge With Caution When Traveling Abroad" because credit-card companies charge 1% to convert foreign-currency expenditures into dollars. In fact, this is the best bargain available to someone traveling abroad.

In contrast to the 1% conversion fee charged by Visa, foreign-currency dealers routinely charge 7% or more to convert U.S. dollars into foreign currency. On top of this, the traveler who converts his dollars into foreign currency before the trip starts will lose interest from the day of conversion. At the end of the trip, any unspent foreign exchange will have to be converted back into dollars, with another commission due. (0980)

3.2.2 Intra-sentential relations

Implicit relations between adjacent clauses in the same sentence not separated by a colon (":") or semi-colon (";") have not been annotated, for example, intra-sentential relations between a main clause and any *free adjunct*. As between adjacent sentences, a variety of relationships can hold between these clauses (Webber and Di Eugenio, 1990):

[13] `Explicit connectives` can, of course, take arguments across a paragraph boundary.

(73) The market for export financing was liberalized in the mid-1980s, forcing the bank to face competition. (0616)

(74) Second, they channel monthly mortgage payments into semiannual payments, reducing the administrative burden on investors. (0029)

(75) Mr. Cathcart says he has had "a lot of fun" at Kidder, adding the crack about his being a "tool-and-die man" never bothered him. (0604)

So in Examples (73) and (74), the event expressed in the free adjunct can be considered a consequence of that expressed in the main clause (which might be annotated with an `Implicit` *so* or *thereby*), while in Example (75) the event expressed in the free adjunct merely follows (as continuation) that expressed in the main clause (which might be annotated with an `Implicit` *then*).

3.2.3 Implicit relations in addition to explicitly expressed relations

We have only annotated implicit relations between adjacent sentences with no `Explicit` connective between them, even though the presence of an `Explicit` connective, in particular a discourse adverbial, in a sentence does not preclude the presence of either another `Explicit` connective relating with the previous text (Example 76) or an `Implicit` connective (Example 77). In both examples, the sentences are related via a causal as well as a conditional relation, with the difference being that the causal relation is expressed with an `Explicit` *because* in Example (76), while the same relation is inferred in Example (77).

(76) If the light is red, stop **because** **otherwise** you'll get a ticket.

(77) If the light is red, stop. **Otherwise** you'll get a ticket.

PDTB-2.0 only annotates multiple `Explicit` relations, as in Example (76), and multiple `Implicit` relations, as in Examples (70-71), but not multiple relations when one of them is explicitly expressed in the text, as in Example (77).

3.2.4 Implicit relations between non-adjacent sentences

Finally, the PDTB does not annotate implicit relations between non-adjacent sentences, even if such a relationship holds. For example, even if the discourse adverbial *then* were removed from Example (78), the event expressed by clause (78d) would still be understood as holding after that expressed by clause (78b). Nevertheless, we neither require nor allow the annotators to annotate the one or more `Implicit` connectives that express the connection holding between clauses (78b) and (78d).

(78) a. John loves Barolo.
b. So he ordered three cases of the '97.
c. But he had to cancel the order
d. because he (then) discovered he was broke.

3.3 Extent of arguments

3.3.1 Sub-sentential arguments

While implicit relations are annotated between adjacent sentences, this does not mean that the arguments of an inferred relation need span complete sentences. As with the `Explicit` connectives, annotators were asked to select only as much of the adjacent sentences as was minimally necessary for the interpretation of the inferred relation. Furthermore, as for `Explicit` connectives, parts of the text that were seen as *relevant* (but not *necessary*) to the interpretation of the relation could be marked as *supplementary* information. For instance, in Example (79), for the inferred exemplification relation, the matrix clause is excluded from `Arg1`, and is marked as `Sup1` - its relevance being due to its containment of the referent of the relative pronoun *when* in `Arg1`.

(79) ($_{Sup1}$ Average maturity was as short as 29 days at the start of this year), *when short-term interest rates were moving steadily upward.* `Implicit` = FOR EXAMPLE **The average seven-day compound yield of the funds reached 9.62% in late April**. (0982)

Parts of the sentence may also be left out without being labeled as supplementary information, when they are not considered relevant to the interpretation of the relation, as for example, the non-restrictive relative clause in the sentence containing `Arg2` in Example (80).

(80) *Meanwhile, the average yield on taxable funds dropped nearly a tenth of a percentage point, the largest drop since midsummer.* `implicit` = IN PARTICULAR **The average seven-day compound yield**, which assumes that dividends are reinvested and that current rates continue for a year, **fell to 8.47%, its lowest since late last year, from 8.55% the week before, according to Donoghue's**. (0982)

Attribution is also a cause for selection of sub-sentential spans, as seen in the sentence containing `Arg1` in Example (81), and both the sentences containing `Arg1` and `Arg2` in Example (82).

(81) *"Lower yields are just reflecting lower short-term interest rates,"* said Brenda Malizia Negus, editor of Money Fund Report. `Implicit` = SINCE **Money funds invest in such things as short-term Treasury securities, commercial paper and certificates of deposit, all of which have been posting lower interest rates since last spring**. (0982)

(82) Ms. Terry did say *the fund's recent performance "illustrates what happens in a leveraged product" when the market doesn't cooperate.* `Implicit` = STILL "**When the market turns around**," she says, "**it will give a nice picture" of how leverage can help performance**. (0983)

3.3.2 Multiple sentence arguments

In addition to selecting sub-sentential clauses, arguments (`Arg1` as well as `Arg2`) can also span over multiple sentences (discontinuously, if necessary) if such an extension is *minimally required* for the interpretation of the relation.[14] For instance, for the inferred exemplification relation in

[14] Extension of arguments to multiple sentences is restricted for `EntRel` and `NoRel` (see Section 3.4).

Example (83), the example of *legal controversies always assuming a symbolic significance far beyond the particular case* is given not just by the sentence following it, but rather by a combination of the three following sentences.

(83) Legal controversies in America have a way of assuming a symbolic significance far exceeding what is involved in the particular case. They speak volumes about the state of our society at a given moment.
It has always been so. Implicit = FOR EXAMPLE **In the 1920s, a young schoolteacaher, John T. Scopes, volunteered to be a guinea pig in a test case sponsored by the American Civil Liberties Union to challenge a ban on the teaching of evolution imposed by the Tennessee Legislature. The result was a world-famous trial exposing profound cultural conflicts in American life between the "smart set," whose spokesman was H.L. Mencken, and the religious fundamentalists, whom Mencken derided as benighted primitives. Few now recall the actual outcome: Scopes was convicted and fined $100, and his conviction was reversed on appeal because the fine was excessive under Tennessee law**. (0946)

Similar scenarios obtain for `Explicit` connectives, except that for `Implicit` connective arguments, the extension to multiple sentences is subject to the strict constraint of *adjacency*. That is, at least some part of the spans selected for `Arg1` and `Arg2` must belong to the pair of adjacent sentences initially identified for annotation of the `Implicit` connective.

Lists, when they span multiple sentences, are also taken to be *minimal*. `Arg1` is extended to include the complete list in Example 84.

(84) All the while, Ms. Bartlett had been busy at her assignment, serene in her sense of self-tilth. *As she put it in a 1987 lecture at the Harvard Graduate School of Design: "I have designed a garden, not knowing the difference between a rhododendron and a tulip." Moreover, she proclaimed that "landscape architects have been going wrong for the last 20 years" in the design of open space. And she further stunned her listeners by revealing her secret garden design method: Commissioning a friend to spend "five or six thousand dollars . . . on books that I ultimately cut up." After that, the layout had been easy. "I've always relied heavily on the grid and found it never to fail."* Implicit = IN ADDITION **Ms. Bartlett told her audience that she absolutely did not believe in compromise or in giving in to the client "because I don't think you can do watered-down versions of things**." (0984)

Example (85) shows an example where multiple sentences are selected for both `Arg1` and `Arg2`, as *minimally* required for `Arg2`, and as a *list* for `Arg1`.

(85) While the model was still on view, *Manhattan Community Board 1 passed a resolution against South Gardens. The Parks Council wrote the BPCA that this "too 'private' . . . exclusive," complex and expensive "enclosed garden . . . belongs in almost any location but the waterfront."* Implicit = SIMILARLY **Lynden B. Miller, the noted public garden designer who restored Central Park's Conservatory Garden, recalls her reaction to the South Gardens model in light of the public garden she was designing for 42nd**

Street's Bryant Park: "Bryant Park, as designed in 1933, failed as a public space, because it made people feel trapped. By removing the hedges and some walls, the Bryant Park Restoration is opening it up. It seems to me the BPCA plan has the potential of making South Gardens a horticultural jail for people and plants." (0984)

3.3.3 Arguments involving parentheticals

Implicit relations between parentheticals and adjacent material to the left and right of the parentheses are annotated slightly differently. An implicit relation can be annotated between a parenthetical sentence and the sentence outside the parentheses that precedes it. However, when annotating an implicit relation between a parenthetical and the sentence that follows it after the parentheses, `Arg1` is (at least) extended to the sentence occurring before the parenthetical. So given a three sentence text containing S1, (S2), and S3, where (S2) is the parenthetical, two relations are marked: one between [S1] as `Arg1` and [(S2)] as `Arg2`, and the other between [S1,(S2)] as `Arg1` and [S3] as `Arg2`.

3.4 Non-insertability of `Implicit` connectives

In many cases, an `Implicit` connective cannot be inserted between adjacent sentences. These have been classified into 3 types: `AltLex`, `EntRel`, and `NoRel`.[15] We describe each of these types below.

3.4.1 `AltLex` (Alternative lexicalization)

These are cases where a discourse relation *is* inferred between adjacent sentences but where providing an `Implicit` connective leads to *redundancy in the expression of the relation*. This is because the relation is *alternatively lexicalized* by some "non-connective expression". Such expressions include (1) those which have two parts, one referring to the relation and another anaphorically to `Arg1`; (2) those which have just one part referring anaphorically to `Arg1`; (3) those which have just one part referring to the relation. Some examples of the first kind are given below. Note that the annotation does not make any further distinctions between different types of `AltLex` expressions. In examples below, the `AltLex` expression is shown in square brackets for clarity.

(86) And she further stunned her listeners by revealing her secret garden design method: *Commissioning a friend to spend "five or six thousand dollars . . . on books that I ultimately cut up."* AltLex **[After that], the layout had been easy.** (0984)

(87) I read the exerpts of Wayne Angell's exchange with a Gosbank representative ("Put the Soviet Economy on Golden Rails," editorial page, Oct. 5) with great interest, since the gold standard is one of my areas of research. Mr. Angell is incorrect when he states that the Soviet Union's large gold reserves would give it "great power to establish credibility." *During the latter part of the 19th century, Russia was on a gold standard and had gold reserves*

[15] Note that in previous work (Prasad *et al.*, 2005), we used different labels for two of these categories, NOCONN-ENT for `EntRel`, and NOCONN for `AltLex`.

representing more than 100% of its outstanding currency, but no one outside Russia used rubles.

The Bank of England, on the other hand, had gold reserves that averaged about 30% of its outstanding currency, and Bank of England notes were accepted throughout the world. AltLex **[The most likely reason for this disparity] is that the Bank of England was a private bank with substantial earning assets, and the common-law rights of creditors to collect claims against the bank were well established in Britain.** (0985)

(88) *Ms. Bartlett's previous work, which earned her an international reputation in the non-horticultural art world, often took gardens as its nominal subject.* AltLex **[Mayhap this metaphorical connection made] the BPC Fine Arts Committee think she had a literal green thumb**. (0984)

Annotation of the arguments of AltLex follows the same guidelines as for arguments of Implicit connectives. That is, they are subject to the *adjacency constraint*, they can be discontinuous, and they must include all and only the amount of text *minimally* required for interpretating the relation.

3.4.2 EntRel (Entity-based coherence)

EntRel captures cases where the implicit relation between adjacent sentences is not between their AO interpretations, but is rather a form of *entity-based coherence* (Knott *et al.*, 2001) in which the same entity is realized in both sentences, either directly (Examples 89-90) or indirectly (Example 91).[16] Note that entity realization here also includes reification of an abstract object (AO) mentioned in the first sentence, such as with the demonstrative *this* in Example (92), and the definite description *the appointments* in Example (93).

(89) *Hale Milgrim, 41 years old, senior vice president, marketing at Elecktra Entertainment Inc., was named president of Capitol Records Inc., a unit of this entertainment concern.* EntRel **Mr. Milgrim succeeds David Berman, who resigned last month**. (0945)

(90) *The purchase price was disclosed in a preliminary prospectus issued in connection with MGM Grand's planned offering of six million common shares.* EntRel **The luxury airline and casino company, 98.6%-owned by investor Kirk Kerkorian and his Tracinda Corp., earlier this month announced its agreements to acquire the properties, but didn't disclose the purchase price**. (0981)

(91) *Last year the public was afforded a preview of Ms. Bartlett's creation in a tablemodel version, at a BPC exhibition.* EntRel **The labels were breathy: "Within its sheltering walls is a microcosm of a thousand years in garden design ... At the core of it all is a love for plants**." (0984)

(92) She has done little more than recycle her standard motifs – trees, water, landscape fragments, rudimentary square houses, circles, triangles, rectangles – and fit them into a grid, as if she

[16] We use the term *indirect realization* as used in Centering Theory (Grosz *et al.*, 1995), to refer to *inferrables*, and more generally, the phenomenon of *bridging*.

were making one of her gridded two-dimensional works for a gallery wall. *But for South Gardens, the grid was to be a 3-D network of masonry or hedge walls with real plants inside them.* EntRel **In a letter to the BPCA, kelly/varnell called this "arbitrary and amateurish."** (0984)

(93) *Ronald J. Taylor, 48, was named chairman of this insurance firm's reinsurance brokerage group and its major unit, G.L. Hodson & Son Inc. Robert G. Hodson, 65, retired as chairman but will remain a consultant. Stephen A. Crane, 44, senior vice president and chief financial and planning officer of the parent, was named president and chief executive of the brokerage group and the unit, succeeding Mr. Taylor.* EntRel **The appointments are effective Nov. 1**. (0948)

EntRel annotations are not associated with any sense, their labels being self-evident of their semantic type. Argument selection for EntRel is subject to the *adjacency constraint*, though the selection can be discontinuous. The "minimality" constraint here is somewhat restricted, in that the selection should be minimal *up to* the level of the sentence. In particular, for EntRel we only identify the minimal set of (complete) sentences that mention the entities reified in the Arg2 sentence. Thus, unlike Explicit, Implicit and AltLex annotations, arguments of the EntRel relation cannot comprise a sub-sentential span, including those obtained by excluding attribution. In Example (94), for instance, the entire sentences are selected as Arg1 and Arg2, even though the "remodeling" and "refurbishing" event entities in Arg1 that are reified and predicated of in Arg2 are embedded as conjoined arguments in the sentential complement, and even though the reification and predication of the same entities in Arg2 should strictly exclude two levels of attribution (see Section 5).

(94) *Proceeds from the offering are expected to be used for remodeling the company's Desert Inn resort in Las Vegas, refurbishing certain aircraft of the MGM Grand Air unit, and to acquire the property for the new resort.* EntRel **The company said it estimates the Desert Inn remodeling will cost about $32 million, and the refurbishment of the three DC-8-62 aircraft, made by McDonnell Douglas Corp., will cost around $24.5 million**. (0981)

Example (95) illustrates an annotation of EntRel where multiple sentence arguments are required. The last sentence only provides an additional predication about the two mentioned ads, but since the antecedent of the referring expression, *both ads*, is "split" across the previous two sentences, both sentences are selected as Arg1 of the EntRel relation.

(95) HOLIDAY ADS: Seagram will run two interactive ads in December magazines promoting its Chivas Regal and Crown Royal brands. *The Chivas ad illustrates – via a series of pullouts – the wild reactions from the pool man, gardener and others if not given Chivas for Christmas. The three-page Crown Royal ad features a black-and-white shot of a boring holiday party – and a set of colorful stickers with which readers can dress it up.* EntRel **Both ads were designed by Omnicom's DDB Needham agency**. (0989)

Supplementary annotations are disallowed for arguments of EntRel. We also do not provide any further annotation within the arguments to identify the entity or entities realized across the arguments:

annotation of explicit or implicit anaphoric relations not associated directly with discourse relations is outside the scope of this project.

3.4.3 NoRel (No relation)

These are cases where no discourse relation or entity-based coherence relation can be inferred between adjacent sentences. Examples (96-98) show cases where the NoRel label was used.

(96) The products already available are cross-connect systems, used instead of mazes of wiring to interconnect other telecommunications equipment. *This cuts down greatly on labor, Mr. Buchner said.* NoRel **To be introduced later are a multiplexer, which will allow several signals to travel along a single optical line; a light-wave system, which carries voice channels; and a network controller, which directs data flow through cross-connect systems**. (1064)

(97) Jacobs Engineering Group Inc. 's Jacobs International unit was selected to design and build a microcomputer-systems manufacturing plant in County Kildare, Ireland, for Intel Corp. *Jacobs is an international engineering and construction concern.* NoRel **Total capital investment at the site could be as much as $400 million, according to Intel**. (1081)

(98) While the model was still on view, Manhattan Community Board 1 passed a resolution against South Gardens. The Parks Council wrote the BPCA that this "too 'private' . . . exclusive," complex and expensive "enclosed garden . . . belongs in almost any location but the waterfront." Lynden B. Miller, the noted public garden designer who restored Central Park's Conservatory Garden, recalls her reaction to the South Gardens model in light of the public garden she was designing for 42nd Street's Bryant Park: "Bryant Park, as designed in 1933, failed as a public space, because it made people feel trapped. *By removing the hedges and some walls, the Bryant Park Restoration is opening it up.* NoRel **It seems to me the BPCA plan has the potential of making South Gardens a horticultural jail for people and plants**." (0984)

As with EntRel annotations, NoRel does not imply that the material in Arg2 is not related to anything: It is just that the PDTB does not annotate implicit relations between non-adjacent sentences (Section 3.2.4), such as the conjunctive relation between the claim in the last sentence of Example (98) and the claim in its antepenultimate sentence.

For Norel annotations, all and only the adjacent sentences are annotated as the arguments. *Supplementary* annotations are disallowed. And obviously, because of the absence of a relation, no sense annotation is recorded.

4 Senses

4.1 Introduction

Senses have been annotated in the form of sense tags for `Explicit` and `Implicit` connectives, and `AltLex` relations. Depending on the context, the content of the arguments and possibly other factors, discourse connectives, just like verbs, can have more than one sense. In such cases, the purpose of sense annotation is to indicate which of these may hold. In all cases, sense tags provide a semantic description of the relation between the arguments of connectives. When the annotators identify more that one simultaneous interpretation, multiple sense tags are provided. However, arguments may also be related to one another in ways that do not have corresponding sense tags. So sense annotation specifies one or more, but not necessarily all, the semantic relations holding between the arguments of the connectives.

In what follows, we give an overview of the set of sense tags used in the PDTB followed by individual descriptions of each tag and examples from the corpus. In Section 4.7, we discuss the connectives *as if*, *even if*, *otherwise*, and *so that* whose sense labelling in PDTB requires additional discussion.

4.2 Hierarchy of senses

The tagset of senses is organized hierarchically (cf. Figure 1). The top level, or *class level*, has four tags representing four major semantic classes:"TEMPORAL", "CONTINGENCY", "COMPARISON" and "EXPANSION". For each class, a second level of *types* is defined to further refine the semantics of the class levels. For example, "CONTINGENCY" has two types "Cause" (relating two situations via a direct cause-effect relation) and "Condition" (relating a hypothetical scenario with its (possible) consequences). A third level of *subtype* specifies the semantic contribution of each argument. For "CONTINGENCY", its "Cause" type has two subtypes – "reason" (which applies when the connective indicates that the situation specified in `Arg2` is interpreted as the cause of the situation specified in `Arg1`, as often with the connective *because*) and "result" (which is used when the connective indicates that the situation described in `Arg2` is interpreted as the result of the situation presented in `Arg1`. A connective typically tagged as "result" is "as a result".

For most types and subtypes, we also provide some hints about their possible formal semantics. In doing so, we do not attempt to represent the internal meaning of `Arg1` and `Arg2`, but simply refer to them as ||Arg1|| and ||Arg2|| respectively. While these hints are meant to be a starting point for the definition of an integrated logical framework able to deal with the semantics of discourse connectives, they can also help annotators in choosing the proper sense tag.

The hierarchical organization of the sense tags serves two purposes. First, it allows the annotations to be more flexible and thus more reliable. This is because the annotators can choose to annotate at a level that is comfortable to them: they are not forced to provide finer semantic descriptions than they are confident about or which the context does not sufficiently disambiguate. Secondly, the hierarchical organization of tags also allows useful inferences at all levels. For example, Section 4.5.3 illustrates a case where neither the text nor annotators' world knowledge has been sufficient to enable them to provide a sense tag at the level of subtype. Instead, they have provided one at the level of type.

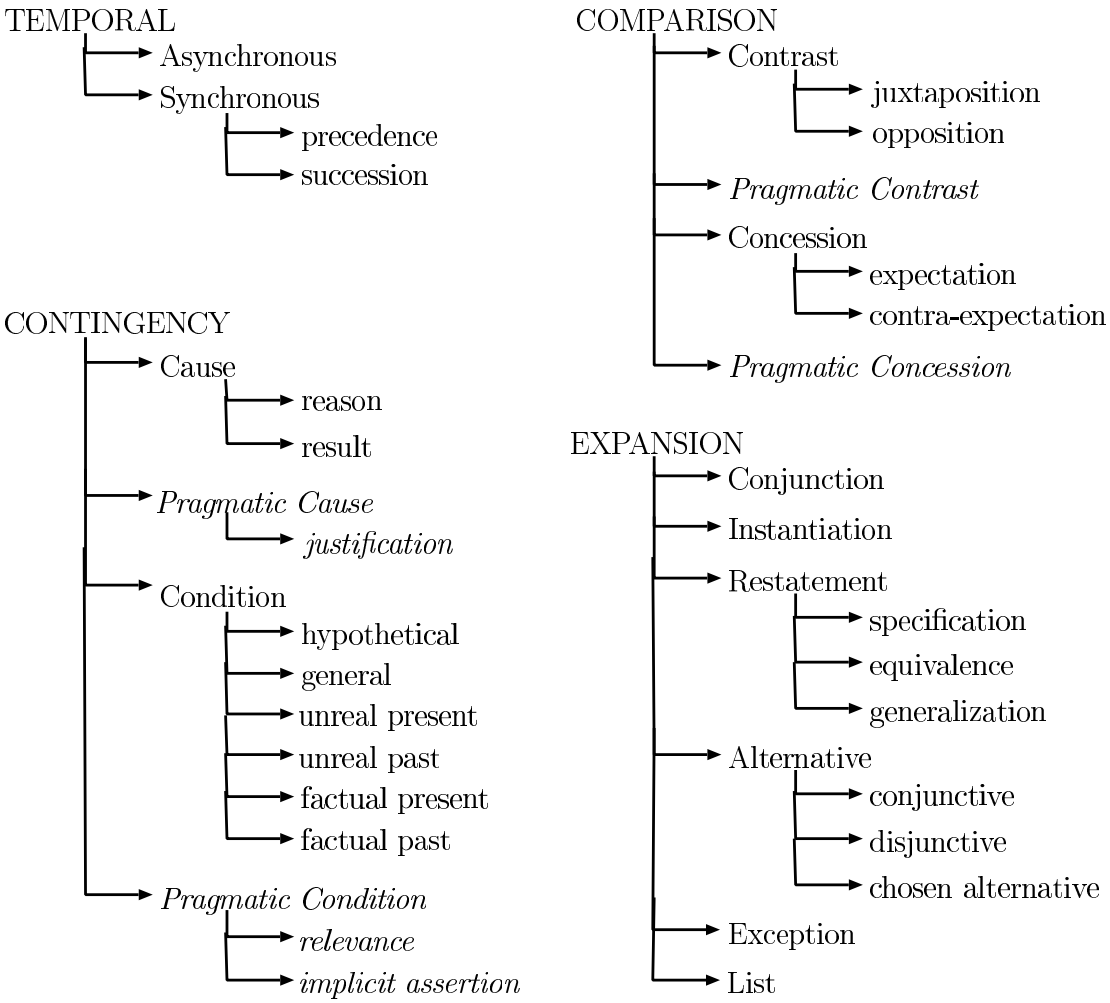

Figure 1: Hierarchy of sense tags

Connectives can also be used to relate the *use* of the arguments of a connective to one another or the use of one argument with the sense of the other. For these *rhetorical* or *pragmatic* uses of connectives, we have defined *pragmatic* sense tags – specifically, "Pragmatic Cause", "Pragmatic Condition", "Pragmatic Contrast" and "Pragmatic Concession".

In the following sections, we provide descriptions of all the class, type and subtype tags used in the annotation of sense in PDTB as well as pragmatic sense tags. Class level tags appear fully capitalized, type level tags start with upper-case and subtype level tags are in lowercase. All sense tags are in quotations marks.

4.3 Class: "TEMPORAL"

The tag "TEMPORAL" is used when the connective indicates that the situations described in the arguments are related temporally. The class level tag "TEMPORAL" does not specify if the situations are temporally ordered or overlapping. Two types are defined for "TEMPORAL": "Asynchronous" (i.e., temporally ordered) and "Synchronous" (i.e., temporally overlapping).

4.3.1 Type: "Asynchronous"

The tag "Asynchronous" is used when the connective indicates that the situations described in the two arguments are temporally ordered. Two subtypes are defined which specify whether it is `Arg1` or `Arg2` that describes an earlier event.

Subtype: "precedence" is used when the connective indicates that the situation in `Arg1` precedes the situation described in `Arg2`, as *before* does in (99).

(99) But a Soviet bank here would be crippled unless Moscow found a way to settle the $188 million debt, *which was lent to the country's short-lived democratic Kerensky government* before **the Communists seized power in 1917**. (TEMPORAL:Asynchronous:precedence) (0035)

Subtype "succession" is used when the connective indicates that the situation described in `Arg1` follows the situation described in `Arg2`, as *after* does in (100).

(100) No matter who owns PS of New Hampshire, after **it emerges from bankruptcy proceedings** *its rates will be among the highest in the nation*, he said. (TEMPORAL:Asynchronous:succession) (0013)

4.3.2 Type: "Synchronous"

The tag "Synchronous" applies when the connective indicates that the situations described in `Arg1` and `Arg2` overlap. The type "Synchronous" does not specify the form of overlap, i.e., whether the two situations started and ended at the same time, whether one was temporally embedded in the other, or whether the two crossed. Typical connectives tagged as "Synchronous" are *while* and *when*, the latter shown in (101).

(101) *Knowing a tasty – and free – meal* when **they eat one**, the executives gave the chefs a standing ovation. (TEMPORAL:Synchrony) (0010)

4.4 Class: "CONTINGENCY"

The class level tag "CONTINGENCY" is used when the connective indicates that one of the situations described in `Arg1` and `Arg2` causally influences the other.

4.4.1 Type: "Cause"

The type "Cause" is used when the connective indicates that the situations described in `Arg1` and `Arg2` are causally influenced and the two are not in a conditional relation. The directionality of causality is not specified at this level: when "Cause" is used in annotation, it means that the annotators could not uniquely specify its directionality. Directionality is specified at the level of subtype: "reason" and "result" specify which situation is the cause and which, the effect.

The rough formal semantics of "Cause" follows Giordano and Schwind (2004) in modelling causality with the binary operator $<$ such that A$<$B models the causal law "A causes B".[17] Here A and B are drawn from the situations described in ||Arg1|| and ||Arg2||. Unless the connective and its arguments are embedded in a matrix that alters their truth value, the situations denoted by A and B and the causal relation between them are all taken to hold.

Subtype: "reason". The type "reason" is used when the connective indicates that the situation described in `Arg2` is the cause and the situation described in `Arg1` is the effect ($||\text{Arg2}|| < ||\text{Arg1}||$), as shown in (102).

(102) *Use of dispersants was approved* when **a test on the third day showed some positive results**, officials said. (CONTINGENCY:Cause:reason) (1347)

Subtype: "result". The type "result" applies when the connective indicates that the situation in `Arg2` is the effect brought about by the situation described in `Arg1`, ($||\text{Arg1}|| < ||\text{Arg2}||$), as shown in (103).

(103) *In addition, its machines are typically easier to operate*, so **customers require less assistance from software**. (CONTINGENCY:Cause:result) (1887)

4.4.2 Type: "Pragmatic Cause"

The tag "Pragmatic Cause" with the subtype label "justification" is used when the connective indicates that `Arg1` expresses a claim and `Arg2` provides justification for this claim, as shown in the use of 'because' in (104). There is no causal influence between the two situations. Epistemic uses of the connective "because" are labelled as "Pragmatic cause:justification". While no instances have been found in the corpus of an `Explicit` or `Implicit` connective in which "Pragmatic cause" holds in the opposite direction (i.e., with `Arg2` expressing the claim and `Arg1` the justification), we allow for this by making "justification" a subtype. However, currently no semantic distinction is made between the type "Pragmatic" and the subtype "justification".

(104) *Mrs Yeargin is lying.* `Implicit =` BECAUSE **They found students in an advanced class a year earlier who said she gave them similar help**. (CONTINGENCY:Pragmatic Cause:justification) (0044)

4.4.3 Type: "Condition"

The type "Condition" is used to describe all subtypes of conditional relations. In addition to causal influence, "Condition" allows some basic inferences about the semantic contribution of the arguments. Specifically, the situation in `Arg2` is taken to be the condition and the situation described in `Arg1` is taken to be the consequence, i.e., the situation that holds when the condition is true. Unlike "Cause", however, the truth value of the arguments of a "Condition" relation cannot be determined independently of the connective.

[17] Logical implication ($\rightarrow$) is used in the rough semantics of "Restatement" (cf. Section 4.6.2).

For this reason, we introduce some branching-time logic operators into our rough description of the semantics of "Condition" subtypes: A, F, and G. A universally quantifies over all possible futures; therefore, $A\beta$ is true iff β is true in all possible futures. F and G are respectively existential and universal quantifiers over instants in a single future: $F\alpha$ is true iff α is true in some instant in a possible future, while $G\alpha$ is true iff α is true in every instant in a possible future.

Subtype: "hypothetical". The semantics for "hypothetical" is $||\text{Arg2}|| < AF||\text{Arg1}||$: if `Arg2` holds true, `Arg1` is caused to hold at some instant in all possible futures. However, `Arg1` can be true in the future independently of `Arg2`.

The condition (`Arg2`) is evaluated in the present and the future. An example tagged as "hypothetical" is given in (105). The verbs in `Arg1` and `Arg2` are usually in present or future tense, except when the conditional is embedded under a report verb in past tense, as shown in (106). In such cases, we map the conditional to its direct form and tag it appropriately. In (106), we assume that the direct form is *Black & Decker will sell two other Emhart operations if it receives the right price.*

(105) Both sides have agreed *that the talks will be most successful* if **negotiators start by focusing on the areas that can be most easily changed**. (CONTINGENCY:Condition:hypothetical) (0082)

(106) In addition, Black & Decker had said *it would sell two other undisclosed Emhart operations* if **it received the right price**. (CONTINGENCY:Condition:hypothetical) (0807)

Subtype: "general". The tag "general" applies if the connective indicates that every time that ||Arg2|| holds true , ||Arg1|| is also caused to be true. Typically, "general" describes either a generic truth about the world or a statement that describes a regular outcome every time the condition holds true. Its semantics is then $AG(||\text{Arg2}|| < ||\text{Arg1}||)$: in all possible futures, it is always the case that ||Arg2|| causes ||Arg1||. The verbs in `Arg1` and `Arg2` are typically in present and future tenses. An example of "general" is shown in (107).

(107) That explains why the number of these wines is expanding so rapidly. But consumers who buy at this level are also more knowledgeable than they were a few years ago. "*They won't buy* if **the quality is not there**," said Cedric Martin of Martin Wine Cellar in New Orleans. (CONTINGENCY:Condition:general) (0071)

The main difference between "hypothetical" and "general" is that, in the former, the causal relation is taken to hold at a single time. For example, (105) says that the talks will be most successful if *now* the negotiators start by focusing on the areas that can be most easily changed. In the future, this may no longer be true: even if the negotiators will start to focus on those areas, the talks may be unsuccessful (i.e., in the future, there may be other factors that affect the performance of the talks).

Subtype: "factual present". The tag "factual present" applies when the connective indicates that `Arg2` is a situation that has either been presented as a fact in the prior discourse or is believed by somebody other than the speaker/writer. "Factual present" is really a special case of the subtype "hypothetical". Besides asserting the condition between the two arguments, it also asserts that ||Arg2|| holds true or is believed by someone to hold true. (If ||Arg2|| indeed holds true, then

||Arg1|| is caused to be true.) We can represent that ||Arg2|| is believed by someone to hold true by means of an epistemic operator $Bel(||\text{Arg2}||)$. Therefore, the semantics for factual present is $||\text{Arg2}|| < AF||\text{Arg1}|| \wedge (||\text{Arg2}|| \vee Bel(||\text{Arg2}||))$. An example of "factual present" is shown in (108).

(108) "I've heard that there is $40 billion taken in nationwide by boiler rooms every year," Mr. McClelland says. "If **that's true**, *Orange County has to be at least 10% of that.*" (CONTINGENCY:Condition:factual present) (1568)

Subtype: "factual past". The tag "factual past" is similar to "factual present" except that in this case `Arg2` describes a situation that is assumed to have taken place at a time in the past. In (109), for example, the speaker expresses in `Arg2` what in the prior discourse is asssumed to have taken place, and in `Arg1`, a consequence that may subsequently occur assuming `Arg2` holds.

(109) "If **they had this much trouble with Chicago & North Western**, *they are going to have an awful time with the rest.*" (CONTINGENCY:Condition:factual past) (1464)

Subtype: "unreal present". The tag "unreal present" applies when the connective indicates that `Arg2` describes a condition that either does not hold at present, e.g., (110) or is considered unlikely to hold e.g., (111). `Arg1` describes what would also hold if `Arg2` were true. The tag "unreal present" represents the semantics of conditional relations also known in the lingustic literature as present *counterfactuals* (Iatridou, 2000). The semantics for "unreal present" is a special case of the semantics for hypothetical. Besides asserting the condition between the two arguments, we also assert that $\sim||\text{Arg2}||$ (meaning ||Arg|| does not hold or is not expected to hold), i.e. $||\text{Arg2}|| < AF||\text{Arg1}|| \wedge \sim||\text{Arg2}||$

(110) *Of course*, if **the film contained dialogue**, *Mr. Lane's Artist would be called a homeless person.* (CONTINGENCY:Condition:unreal present) (0039)

(111) I'm not saying advertising revenue isn't important," she says, "*but I couldn't sleep at night*" if **the magazine bowed to a company because they once took out an ad**. (CONTINGENCY:Condition:unreal present) (0062)

Subtype: "unreal past". The subtype "unreal past" applies when the connective indicates that `Arg2` describes a situation that did not occur in the past and `Arg1` expresses what the consequence would have been if it had. An example is shown in (112). It is inferred from the semantics of this subtype of "Condition" that the situations described in `Arg1` and `Arg2` did not hold.

(112) "If **I had come into Friday on margin or with very little cash in the portfolios**, *I would not do any buying.* (CONTINGENCY:Condition:unreal past) (2376)

4.4.4 Type: "Pragmatic Condition"

The tag "pragmatic condition" is used for instances of conditional constructions whose interpretation deviates from that of the semantics of "Condition". Specifically, these are cases of `Explicit` *if* tokens with `Arg1` and `Arg2` not being causally related. In all cases, `Arg1` holds true independently of `Arg2`.

Subtype: "relevance". The conditional clause in the "relevance" conditional (`Arg2`) provides the context in which the description of the situation in `Arg1` is relevant. A frequently cited example for this type of conditional is (113) and a corpus example is given in (114). There is no causal relation between the two arguments.

(113) If **you are thirsty**, *there's beer in the fridge.*

(114) If **anyone has difficulty imagining a world in which history went merrily on without us**, *Mr. Gould sketches several.* (CONTINGENCY:Pragmatic condition:relevance) (1158)

Subtype: "implicit assertion". The tag "implicit assertion" applies in special rhetorical uses of if-constructions when the intepretation of the conditional construction is an implicit assertion. In (115), for example, `Arg1`, *O' Connor is your man* is not a consequent state that will result if the condition expressed in `Arg2` holds true. Instead, the conditional construction in this case implicitly asserts that O'Connor will keep the crime rates high.

(115) In 1966, on route to a re-election rout of Democrat Frank O'Connor, GOP Gov. Nelson Rockefeller of New York appeared in person saying, "If **you want to keep the crime rates high**, *O'Connor is your man.*"
(CONTINGENCY:Pragmatic Condition:implicit assertion) (0041)

4.5 Class: COMPARISON

The class tag "COMPARISON" applies when the connective indicates that a discourse relation is established between `Arg1` and `Arg2` in order to highlight prominent differences between the two situations. Semantically, the truth of both arguments is independent of the connective or the established relation. "COMPARISON" has two types to further specify its semantics. In some cases, `Arg1` and `Arg2` share a predicate or a property and the difference is highlighted with respect to the values assigned to this property. This interpretation is tagged with the type "Contrast". There are also cases in which the highlighted differences are related to expectations raised by one argument which are then denied by the other. This intepretation is tagged with the type "Concession".

4.5.1 Type: "Contrast"

"Contrast" applies when the connective indicates that `Arg1` and `Arg2` share a predicate or property and a difference is highlighted with respect to the values assigned to the shared property. In "Contrast", neither argument describes a situation that is asserted on the basis of the other one. In this sense, there is no directionality in the interpretation of the arguments. This is an important difference between the interpretation of "Contrast" and "Concession". Two subtypes of "Contrast" are defined: "juxtaposition" and "opposition".

Subtype: "juxtaposition". The subtype "juxtaposition" applies when the connective indicates that the values assigned to some shared property are taken to be alternatives (e.g., *John paid $5 but Mary paid $10.*) More than one shared predicate or property may be juxtaposed. In (116), the

shared predicate *rose* or *jumped* takes two different values (69% and 85%) and the shared predicate *rose to X amount* applies to two entities (*the operating revenue* and *the net interest bill*). When the intended juxtaposition is not clear, the higher level tag "Contrast" is annotated.

(116) *Operating revenue rose 69% to* A$*8.48 billion from* A$*5.01 billion.* But **the net interest bill jumped 85% to A$686.7 million from A$371.1 million**.
(COMPARISON:Contrast:juxtaposition) (1449)

Subtype: "opposition". The subtype "opposition" applies when the connective indicates that the values assigned to some shared property are the extremes of a gradable scale, e.g., tall-short, accept-reject etc.

Note that the notion of gradable scale used in distinguishing "opposition" from "juxtaposition" strongly depends on the context where the sentence is uttered. For example, consider the pair black-white. These two concepts are usually taken to be antonyms. Therefore, it seems that whenever `Arg1` assigns 'black' and `Arg2` assigns 'white' to a shared property (e.g. *Mary is black whereas John is white*), the discourse connective has to be labelled as "opposition". However, in many contexts 'black' and 'white' are just two of the colors that may be assigned to the shared property (e.g., take the sentence *Mary bought a black hat whereas John bought a white one* uttered in a shop that sells red, yellow and blue hats as well). In such cases, they are not antonyms, and the connective is labelled as "juxtaposition".

(117) *Most bond prices fell on concerns about this week's new supply and disappointment that stock prices didn't stage a sharp decline.* **Junk bond prices moved higher**, however.
(COMPARISON:Contrast:opposition) (1464)

4.5.2 Type: "Pragmatic Contrast"

The tag "Pragmatic Contrast" applies when the connective indicates a contrast between one of the arguments and an *inference* that can be drawn from the other, in many cases at the speech act level: The contrast is not between the situations described in `Arg1` and `Arg2`. In (118), for example, the contrast is between `Arg1` and the inference that quantity isn't the only thing that needs to be explained with respect to producers now creating appealing wines: Quality needs to be explained as well, cf. `Arg2`.

(118) "It's just sort of a one-upsmanship thing with some people," added Larry Shapiro. "They like to talk about having the new Red Rock Terrace one of Diamond Creek's Cabernets or the Dunn 1985 Cabernet, or the Petrus. Producers have seen this market opening up and they're now creating wines that appeal to these people." *That explains why the number of these wines is expanding so rapidly.* But **consumers who buy at this level are also more knowledgeable than they were a few years ago**. (COMPARISON:Pragmatic Contrast) (0071)

4.5.3 Type: "Concession"

The type "Concession" applies when the connective indicates that one of the arguments describes a situation A which **causes** C, while the other asserts (or implies) ¬C. Alternatively, one argument denotes a fact that triggers a set of potential consequences, while the other denies one or more of them. Formally: A<C ∧ B→¬C, where A and B are drawn from ||Arg1|| and ||Arg2||. (¬C may be the same as B, where B→B is always true.)

Two "Concession" subtypes are defined in terms of the argument creating an expectation and the one denying it. Specifically, when `Arg2` creates an expectation that `Arg1` denies **(A=||Arg2|| and B=||Arg1||)**, it is tagged as "expectation", shown in (119). When `Arg1` creates an expectation that `Arg2` denies **(A=||Arg1|| and B=||Arg2||)**, it is tagged as "contra-expectation", shown in (120).

(119) Although **the purchasing managers' index continues to indicate a slowing economy**, *it isn't signaling an imminent recession*, said Robert Bretz, chairman of the association's survey committee and director of materials management at Pitney Bowes Inc., Stamford, Conn. (COMPARISON:Concession:expectation) (0036)

(120) *The Texas oilman has acquired a 26.2% stake valued at more than $1.2 billion in an automotive-lighting company, Koito Manufacturing Co.* But **he has failed to gain any influence at the company**. (COMPARISON:Concession:contra-expectation) (0082)

(121) *Besides, to a large extent, Mr. Jones may already be getting what he wants out of the team*, even though **it keeps losing**. (COMPARISON:Concession) (1411)

Instances have been found in the PDTB which are ambiguous between "expectation" and "contra-expectation", where the context or the annotators' world knowledge is not sufficient to specify the subtype, as in (121). Such cases are tagged as "Concession".

4.6 Class: "EXPANSION"

The class "EXPANSION" covers those relations which expand the discourse and move its narrative or exposition forward. Here we describe its subtypes.

4.6.1 Type: "Instantiation"

The tag "Instantiation" is used when the connective indicates that `Arg1` evokes a set and `Arg2` describes it in further detail. It may be a set of events (122), a set of reasons, or a generic set of events, behaviors, attitudes, etc. Typical connectives often tagged as "Instantiation" are *for example*, *for instance* and *specifically*.

(122) He says *he spent $300 million on his art business this year.* `Implicit =` IN PARTICULAR **A week ago, his gallery racked up a $23 million tab at a Sotheby's auction in New York buying seven works, including a Picasso**. (EXPANSION:Instantiation) (0800)

The rough semantics for "Instantiation" involves (1) both arguments holding – ie, ||Arg1|| ∧ ||Arg2|| – and (2) following (Forbes-Riley *et al.*, 2006), a relation holding between ||Arg1|| and ||Arg2|| of

the form *exemplify'*(||Arg2||, $\lambda x.x \in g(||\text{Arg1}||)$), where g is a function that "extracts" the set of events, reasons, behaviours, etc. from the semantics of `Arg1`, and x is a variable ranging over them. *exemplify'* asserts that ||Arg2|| further describes one element in the extracted set.

4.6.2 Type: "Restatement"

A connective is marked as "Restatement" when it indicates that the semantics of `Arg2` restates the semantics of `Arg1`. It is inferred that the situations described in `Arg1` and `Arg2` hold true at the same time. The subtypes "specification", "generalization", and "equivalence" further specify the ways in which `Arg2` restates `Arg1`: ||Arg1||$\rightarrow$||Arg2|| in the case of generalization, ||Arg1||$\leftarrow$||Arg2|| in the case of specification, and ||Arg1||$\leftrightarrow$||Arg2|| in the case of equivalence, where $\rightarrow$ indicates logical implication.

Subtype: "specification". "Specification" applies when `Arg2` describes the situation described in `Arg1` in more detail, as in (123) and (124). Typical connectives for "specification" are *specifically*, *indeed* and *in fact*.

(123) A Lorillard spokewoman said, "*This is an old story.* `Implicit =` IN FACT **We're talking about years ago before anyone heard of asbestos having any questionable properties**." (EXPANSION:Restatement:specification) (0003)

(124) *An enormous turtle has succeeded where the government has failed*: `Implicit =` SPECIFICALLY **He has made speaking Filipino respectable**.
(EXPANSION:Restatement:specification) (0804)

Subtype: "generalization". "Generalization" applies when the connective indicates that `Arg2` summarizes `Arg1`, or in some cases expresses a conclusion based on `Arg1`. An example of "generalization" is given in (125). Typical connectives for "generalization" are *in sum*, *overall*, *finally*, etc.

(125) *If the contract is as successful as some expect, it may do much to restore confidence in futures trading in Hong Kong.* `Implicit =` IN OTHER WORDS. "**The contract is definitely important to the exchange**," says Robert Gilmore, executive director of the Securities and Futures Commission. (EXPANSION:Restatement:generalization) (0700)

Subtype: "equivalence". "Equivalence" applies when the connective indicates that `Arg1` and `Arg2` describe the same situation from different perspectives, as in (126), where the two arguments highlight two different aspects of the same situation.

(126) Chairman Krebs says *the California pension fund is getting a bargain price that wouldn't have been offered to others.* In other words: **The real estate has a higher value than the pending deal suggests**. (EXPANSION:Restatement:equivalence) (0331)

Whether a relation is a case of "specification" or "equivalence" depends on the `Implicit` connective. In (127), the speaker is taken to be pointing to one of possible things that could be done to avoid gambling too far. In (128), the speaker is taken to be explaining what he or she means by not gambling too far.

(127) *I never gamble too far.* Implicit = IN PARTICULAR. **I quit after one try**.

(128) *I never gamble too far.* Implicit = IN OTHER WORDS. **I quit after one try**.

The Type level tag "Restatement" is used when more than on subtype interpretation is possible, as in (129), where Arg2 can be interpreted as denoting what *he* said, or it can be interepreted as providing the same information from a different point of view, namely the speaker's own words.

(129) He said the assets to be sold would be "non-insurance" assets, including a beer company and a real estate firm, *and wouldn't include any pieces of Farmers.* Implicit = IN OTHER WORDS "**We won't put any burden on Farmers**," he said. (EXPANSION:Restatement) (2403)

4.6.3 Type: "Alternative"

The type "Alternative" applies when the connective indicates that its two arguments denote alternative situations. It has three subtypes: "conjunctive", "disjunctive" and "chosen alternative".

Subtype: "conjunctive". The "conjunctive" subtype is used when the connective indicates that both alternatives hold or are possible (||Arg1|| $\wedge$ ||Arg2||), as in (130), which specifies two options that investors are encouraged to exercise.

(130) Today's Fidelity ad goes a step further, encouraging investors *to stay in the market* or **even to plunge in with Fidelity**. (EXPANSION:Alternative:conjunctive) (2201)

Subtype: "disjunctive". The "disjunctive" subtype is used when two situations are evoked in the discourse but only one of them holds. In (131), for example, the alternatives are *lock in leases* and *buy now*: One cannot do both simultaneously. The semantics of "disjunctive" is ||Arg1|| *xor* ||Arg2||, where A *xor* B $\equiv$ ((A $\vee$ B) $\wedge$ (A$\rightarrow \neg$B) $\wedge$ (B$\rightarrow \neg$A)).

(131) *Those looking for real-estate bargains in distressed metropolitan areas should lock in leases* or **buy now**. (EXPANSION:Alternative:disjunctive) (2444)

Subtype: "chosen alternative". The "chosen alternative" subtype is used when the connective indicates that two alternatives are evoked in the discourse but only one is taken, as with the connective *instead* shown in (132). The semantics is ||Arg1|| *xor* (||Arg2|| $\wedge \neg$||Arg1||), from which ||Arg2|| can be inferred. [18]

(132) Under current rules, even when a network fares well with a 100%-owned series – ABC, for example, made a killing in broadcasting its popular crime/comedy "Moonlighting" —

[18] This subtype illustrates a feature of the *minimality principle* – that one may have to distinguish between the span which *licences* the use of a connective to link to a particular argument and the span from which the interpretation of that argument derives. Sometimes they are the same, sometimes different. And that interpretion may involve inference. So, for example, while in "I'm allergic to peas. Instead I'll eat beans." the span licensing Arg1 and the span from which the interpretation of Arg1 derives are the same – ie, "I'm allergic to peas", the relevant interpretation of Arg1 (||Arg1||) is "I eat peas" – ie, instead of me eating peas, I'll eat beans. As noted, $\neg$||Arg1|| holds. *Instead* and its annotation are discussed at greater length in (Webber *et al.*, 2005; Miltsakaki *et al.*, 2003).

it isn't allowed *to share in the continuing proceeds when the reruns are sold to local stations.* Instead, **ABC will have to sell off the rights for a one-time fee**. (EXPANSION:Alternative:chosen alternative) (2451)

4.6.4 Type: "Exception"

The type "Exception" applies when the connective indicates that `Arg2` specifies an exception to the generalization specified by `Arg1`, as in (133). In other words, `Arg1` is false because `Arg2` is true, but if `Arg2` were false, `Arg1` would be true. The semantics of "Exception" is: $\neg||\text{Arg1}|| \wedge ||\text{Arg2}|| \wedge \neg||\text{Arg2}|| \rightarrow ||\text{Arg1}||$.

(133) *Boston Co. officials declined to comment on Moody's action on the unit's financial performance this year* except **to deny a published report that outside accountants had discovered evidence of significant accounting errors in the first three quarters' results**. (EXPANSION:Exception) (1103)

4.6.5 Type: "Conjunction"

The Type "Conjunction" is used when the connective indicates that the situation described in `Arg2` provides additional, discourse new, information that is related to the situation described in `Arg1`, but is not related to `Arg1` in any of the ways described for other types of "EXPANSION". (That is, the rough semantics of "Conjunction" is simply $||\text{Arg1}|| \wedge ||\text{Arg2}||$.) An example of "Conjunction" is shown in (134). Typical connectives for "Conjunction" are *also*, *in addition*, *additionally*, *further*, etc.

(134) *Food prices are expected to be unchanged, but energy costs jumped as much as 4%, said Gary Ciminero, economist at Fleet/Norstar Financial Group.* **He** also **says he thinks "core inflation,"** which excludes the volatile food and energy prices, **was strong last month**. (EXPANSION:Conjunction) (2400)

4.6.6 Type: "List"

The Type "List" applies when `Arg1` and `Arg2` are members of a list, defined in the prior discourse. "List" does not require the situations specified in `Arg1` and `Arg2` to be directly related. In Example (135), the list defined roughly as *what make besuboru unrecognizable* has as two of its members the content of `Arg1` and `Arg2`.

(135) But other than the fact that besuboru is played with a ball and a bat, it's unrecognizable: *Fans politely return foul balls to stadium ushers*; `Implicit =` AND **the strike zone expands depending on the size of the hitter**; (EXPANSION:List) (0037)

4.7 Notes on a few connectives

There are a few cases where a sense tag used in the PDTB is idiosyncratic to a particular connective.

4.7.1 Connective: *As if*

The semantics of the connective *as if* expresses a similarity between the situation described in Arg1 and the situation described in Arg2. Although none of the sense tags we have defined expresses similarity, we felt there were too few tokens of "as if" (ie, 16 tokens of the `Explicit` connective, and no tokens of `Implicit` "as if") to create special sense tags for it. Rather, we chose to use existing labels. Tokens of "as if" in the corpus have one of two interpretations: concession and manner. The former was annotated using the Concession:contra-expectation label, as in (136). (Such cases involve the negation of Arg2.) In the manner sense of "as if", Arg2 expresses a similarity to the manner in which Arg1 is performed. While the combination of connective plus Arg2 further specifies Arg1, the sense tag "specification" is not appropriate because the event described in Arg1 does not entail the situation in Arg2 (cf. Section 4.6.2). In (137), for example, *shivering* does not entail that the temperature is 20 below zero.

While it is possible that these cases of "as if" should not be taken as expressing a *discourse* relation at all, we have nevertheless kept these annotations in the corpus and labelled all manner interpretations of "as if" with the class label "EXPANSION".

(136) As if **he were still in his old job**, *Mr. Wright*, by resigning with his title instead of being forced from his job, *by law enjoys a $120,000 annual office expense allowance, three paid staffers, up to $67,000 for stationery and telephones and continued use of the franking privilege.* (COMPARISON:Concession:contra-expectation) (0909)

(137) When I realized it was over, I went and stood out in front of the house, waiting and praying for Merrill to come home, *shivering* as if **it were 20 below zero** until he got there. Never in my life have I been so frightened. (EXPANSION) (1778)

4.7.2 Connective: *Even if*

In PDTB the connective *even if* has been sense-tagged as "Concession". `Arg2` of *even if* creates an expectation that is denied in `Arg1`. Idiosyncratic to *even if* is that the situation described in `Arg2` need not hold, whereas it does in other cases of "Concession".

(138) Even if **the gross national product is either flat or in the growth range of 2% to 2.5%**, *"we can handle that,"* Mr. Marcus said. (COMPARISON:Concession:expectation) (0973)

4.7.3 Connective: *Otherwise*

The connective *otherwise* is ambiguous between the two senses "disjunctive alternative", as in (139), and "exception", as in (140).

(139) *Consumers will be able to switch on their HDTV sets and get all the viewing benefits the high-tech medium offers.* Otherwise, **they'd be watching programs that are no different in quality from what they currently view on color TVs**.
(EXPANSION:Alternative:disjunctive) (1386)

(140) *Twenty-five years ago the poet Richard Wilbur modernized the 17th century comedy merely by avoiding "the zounds sort of thing"* as he wrote in his introduction. Otherwise, **the scene remained Celimene's house in 1666**. (EXPANSION:Exception) (1936)

The latter is idiosyncratic to *otherwise* in that while "exception" is defined such that `Arg2` is true, while `Arg1` would be true if `Arg2` were false, here it is the reverse: `Arg1` is true and `Arg2` would be true if `Arg1` were false.

4.7.4 Connectives: *Or* and *when*

There are cases of *or* (141) and *when* (142), which resemble rhetorical uses of *if* labelled as "implicit assertion" (cf. Section 4.4.4). They have been sense tagged as such, even though they are not associated with *if*.

(141) *If you'd really rather have a Buick, don't leave home without the American Express card.* Or **so the slogan might go**. (CONTINGENCY:Pragmatic Condition:implicit assertion) (0116)

(142) *He's right about his subcommittee's responsibilities* when **it comes to obtaining information from prior HUD officials**. (CONTINGENCY:Pragmatic Condition:implicit assertion) (2377)

4.7.5 Connective: *So that*

Discourse connectives that express purpose (eg, *so that*) have been labelled with the sense tag "CONTINGENCY:Cause:result" as shown in (143). `Arg2` of *so that* expresses the situation that is expected to hold as the result of `Arg1`. Idiosyncratic to purpose connectives tagged in this way is that the situation specified in `Arg2` may or may not hold true at a subsequent time, even if `Arg1` does.

(143) Northeast said it would refile its request and still hopes for *an expedited review by the FERC* so that **it could complete the purchase by next summer if its bid is the one approved by the bankruptcy court**. (CONTINGENCY:Cause:result) (0013)

5 Attribution

5.1 Introduction

The relation of *attribution* is a relation of "ownership" between abstract objects and individuals or agents. That is, attribution has to do with ascribing beliefs and assertions expressed in text to the agent(s) holding or making them (Riloff and Wiebe, 2003; Wiebe *et al.*, 2004, 2005). Since we take discourse connectives to convey semantic predicate-argument relations between abstract objects, one can distinguish a variety of cases depending on the attribution of the discourse relation or its arguments. For example, a discourse relation may hold either between the attributions (and the agents of attributions) themselves or only between the abstract object arguments of the attribution, as shown below:[19]

(144) When **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.* (0267)

(145) *Advocates said* *the 90-cent-an-hour rise, to $4.25 an hour by April 1991, is too small for the working poor*, while **opponents argued** **that the increase will still hurt small business and cost many thousands of jobs**. (0098)

In Example (144), the temporal relation denoted by *when* is expressed between the eventuality of *Mr. Green winning the verdict* and *the Judge giving him an additional award*. In Example (145), on the other hand, the contrastive relation denoted by *while* holds between the agent arguments of the attribution relation, which means that the attribution relation is part of the contrast as well. (In all examples in this section, the text spans corresponding to the attribution phrase are shown boxed.)

Abstract object arguments of attributions can be discourse relations as well, as seen in Example (146), where the temporal relation between the two arguments is also being quoted and is thus attributed to an individual other than the writer of the text.

(146) "When **the airline information came through**, *it cracked every model we had for the marketplace*," said a managing director at one of the largest program-trading firms. (2300)

In addition to `Explicit` connectives, attribution in the PDTB is also marked for `Implicit` connectives and their arguments. `Implicit` connectives express discourse relations that the writer intends for the reader to infer. As with `Explicit` connectives, implicit relations intended by the writer are distinguished from those intended by some other agent or speaker that the writer has introduced. For example, while the implicit relation in Example (147) is attributed to the writer, in Example (148), both `Arg1` and `Arg2` have been expressed by another speaker whose speech is being quoted: in this case, the implicit relation is attributed to the other speaker.[20]

[19] We note that while some attribution spans can be identified clearly as the *reporting frames* of Huddleston and Pullum (2002), others are less clearly categorized this way, sometimes appearing as, for example, adverbial phrases, and sometimes not appearing at all (when they have to be inferred anaphorically from the prior context).

[20] Attribution is also annotated for `AltLex` relations, but not for `EntRel` and `NoRel`, since the latter do not indicate the presence of discourse relations.

(147) *The gruff financier recently started socializing in upper-class circles.* Implicit = FOR EXAMPLE Although he says he wasn't keen on going, **last year he attended a New York gala where his daughter made her debut**. (0800)

(148) *"We've been opposed to" index arbitrage "for a long time,"*
said Stephen B. Timbers, chief investment officer at Kemper, which manages $56 billion, including $8 billion of stocks. Implicit = BECAUSE "**Index arbitrage doesn't work, and it scares natural buyers" of stock**. (1000)

The annotation scheme isolates four key properties of attribution, which are annotated as features:

(a) *Source*, which distinguishes between different types of agents (Section 5.2);

(b) *Type*, which encodes the nature of the relationship between agents and AOs, thereby reflecting their factuality (Section 5.3);

(c) *Scopal polarity*, which is marked when surface negated attribution reverses the polarity of the attributed AO (Section 5.4);

(d) *Determinacy*, which signals a context that cancels what would otherwise be an entailment of attribution (Section 5.5).

In addition, to further facilitate the task of identifying attribution, the scheme also annotates the *text span* signaling attribution (Section 5.6), with the goal of highlighting the textual anchors of the features mentioned above. (In what follows, attribution feature values assigned to examples are shown below each example; REL stands for discourse relation; and, as mentioned above, attribution text spans are shown boxed.)

Appendix G and Appendix H give the distribution of distinct feature combinations found for attribution per relation, for `Explicit` connectives, and `Implicit` connectives and `AltLex` relations, respectively.

5.2 Source

The *source* feature distinguishes between:

(a) the writer of the text ("Wr"),

(b) some specific agent introduced in the text ("Ot" for other),

(c) some arbitrary ("Arb") individual(s) indicated via a non-specific reference in the text.

In addition, since attribution can have scope over an entire relation, arguments can be annotated with a fourth value "Inh", to indicate that their source value is inherited from the relation.

Given this scheme for source, there are broadly two possibilities. In the first case, a relation and both its arguments are attributed to the same source, either the writer, as in (149), or some other agent (here, Bill Biedermann), as in (150).

(149) Since **the British auto maker became a takeover target last month**, its ADRs have jumped about 78%. (0048)

	REL	Arg1	Arg2
[Source]	Wr	Inh	Inh

(150) "*The public is buying the market* when **in reality there is plenty of grain to be shipped**," said Bill Biedermann, Allendale Inc. director. (0192)

	REL	Arg1	Arg2
[Source]	Ot	Inh	Inh

As Example (149) shows, text spans for implicit writer attributions (corresponding to implicit speech acts such as *"I write"*, or *"I say"*) are not marked and imply writer attribution by default.[21]

In the second case, one or both arguments have a different source from the relation. In (151), for example, the relation and `Arg2` are attributed to the writer, whereas `Arg1` is attributed to another agent (here, Mr. Green). On the other hand, in (152) and (153), the relation and `Arg1` are attributed to the writer, whereas `Arg2` is attributed to another agent.

(151) When **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.* (0267)

	REL	Arg1	Arg2
[Source]	Wr	Ot	Inh

(152) *Factory orders and construction outlays were largely flat in December* while purchasing agents said **manufacturing shrank further in October**. (0178)

	REL	Arg1	Arg2
[Source]	Wr	Inh	Ot

(153) *There, on one of his first shopping trips, Mr. Paul picked up several paintings at stunning prices.* He paid $2.2 million, for instance, for a still life by Jan Jansz. den Uyl that was expected to fetch perhaps $700,000. The price paid was a record for the artist. (...) Afterward, Mr. Paul is said by Mr. Guterman **to have phoned Mr. Guterman, the New York developer selling the collection, and gloated**. (2113)

	REL	Arg1	Arg2
[Source]	Wr	Inh	Ot

[21] It is also possible for an "Ot" attribution to be implicit for a relation or argument. These, however, are inferred from some explicit occurrence of the source in the prior text, and their attribution spans are marked extra-sententially (see Section 5.6).

Example (154) shows an example of a non-specific "Arb" source indicated by an agentless passivized attribution on `Arg2` of the relation. Note that passivized attributions can also be associated with a specific source when the agent is explicit, as shown in (153), where the explicit agent is Mr. Guterman.[22] "Arb" sources are also identified by the occurrences of adverbs like *reportedly*, *allegedly*, etc., as in Example (155).

(154) Although [index arbitrage is said] **to add liquidity to markets**, [John Bachmann, ... says] *too much liquidity isn't a good thing.* (0742)

	REL	Arg1	Arg2
[**Source**]	Wr	Ot	Arb

(155) *East Germans rallied* as **officials** [**reportedly**] **sought Honecker's ouster**. (2278)

	REL	Arg1	Arg2
[**Source**]	Wr	Inh	Arb

When "Ot" is used to refer to a specific individual as the source, no further annotation is provided to indicate who the "Ot" agent in the text is. Furthermore, as shown in Examples (156-157), multiple "Ot" sources within the same relation do not indicate whether or not they refer to the same or different agents. This is because of our assumption that the text span annotations for attribution, together with an independent mechanism for named entity recognition and anaphora resolution, can be effectively exploited to identify and disambiguate the appropriate references.

(156) *Suppression of the book*, [Judge Oakes observed], *would operate as a prior restraint and thus involve the First Amendment.*
Moreover, [and here Judge Oakes went to the heart of the question], **"Responsible biographers and historians constantly use primary sources, letters, diaries, and memoranda."** (0944)

	REL	Arg1	Arg2
[**Source**]	Wr	Ot	Ot

(157) *The judge was considered imperious, abrasive and ambitious*,
[those who practiced before him say]...Yet, **despite the judge's imperial bearing, no one ever had reason to suspect possible wrongdoing**,
[says John Bognato, president of Cambria County's bar association]. (0267)

	REL	Arg1	Arg2
[**Source**]	Wr	Ot	Ot

[22] In passivized attributions (e.g., in Examples (153) and (154)), the subject of the infinitive raised to the position of main clause subject is included in the attribution text span. This is due to the convention of including in the attribution span all non-clausal complements and modifiers of the attribution predicate (Section 5.6).

5.3 Type

The *type* feature signifies the nature of the relation between an agent and an AO, leading to different inferences about the degree of factuality of the AO. We start by making the well-known disinction of AOs into four sub-types: *assertion propositions*, *belief propositions*, *facts* and *eventualities*.[23] This initial distinction is significant since it corresponds, in part, to the types of attribution relations and the verbs that convey them, and simultaneously allows for a semantic compositional approach to the annotation and recognition of factuality.[24]

5.3.1 Assertion proposition AOs and belief propositions AOs

Proposition AOs involve attribution to an agent of his/her commitment towards the truth of a proposition. A further distinction captures differences in the degree of that commitment, by distinguishing between "assertions" and "beliefs".

Assertion proposition AOs are associated with a communication type of attribution ("Comm" for short), conveyed by standard verbs of communication (Levin, 1993) such as *say*, *mention*, *claim*, *argue*, *explain* etc. In Example (158), the attribution on `Arg1` takes the value "Comm" for *type*. Implicit writer attributions, as with the relation in Example (158), also take the default value "Comm". Note that when an argument's attribution source is not inherited (as with `Arg1` in this example) it takes its own independent value for *type*. This example thus conveys that there are two different attributions expressed within the discourse relation, one for the relation and the other for one of its arguments, and that both involve propositional assertions.

(158) When **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.* (0267)

	REL	Arg1	Arg2
[Source]	Wr	Ot	Inh
[Type]	Comm	Comm	Null

In the absence of an independent occurrence of attribution on an argument, as for `Arg2` of Example (158), a "Null" value for the *type* on the argument means that it needs to be derived by independent (here, undefined) considerations under the scope of the relation. Note that unlike the "Inh" value of the *source* feature, "Null" does not indicate inheritance. In a subordinate clause, for example, while the relation denoted by the subordinating conjunction may be asserted, the clause content itself may be "presupposed", as seems to be the case in (158). However, we found these differences difficult to determine at times, and consequently leave this undefined in the scheme.

Belief proposition AOs are associated with a "belief" type of attribution, conveyed by propositional attitude verbs (Hintikka, 1971) such as *believe*, *think*, *expect*, *suppose*, *imagine*, etc. This type of attribution is thus called "PAtt" for short. An example of a belief attribution is given in (159).

[23] This corresponds roughly to the top-level tier in the AO hierarchy of Asher (1993).

[24] Note that discourse relations are also taken to denote a special class of propositions, called *relational propositions* (Mann and Thompson, 1988) and are themselves treated as abstract objects in the PDTB (Prasad *et al.*, 2005).

(159) Mr. Marcus believes *spot steel prices will continue to fall through early 1990* **and** then **reverse themselves**. (0336)

	REL	Arg1	Arg2
[**Source**]	Ot	Inh	Inh
[**Type**]	PAtt	Null	Null

5.3.2 Fact AOs

Facts AOs involve attribution to an agent of an evaluation towards or knowledge of a proposition whose truth is taken for granted (i.e., presupposed). Fact AOs are associated with a "factive" type of attribution ("Ftv" for short), conveyed by "factive" and "semi-factive verbs" (Kiparsky and Kiparsky, 1971; Karttunen, 1971) such as *regret*, *forget*, *remember*, *know*, *see*, *hear*, etc. An example of a factive attribution is given in (160). However, this class does not distinguish between the true factives and semi-factives, the former involving an attitude/evaluation towards a fact, and the latter involving knowledge of a fact.

(160) The other side, he argues knows *Giuliani has always been pro-choice*, even though **he has personal reservations**. (0041)

	REL	Arg1	Arg2
[**Source**]	Ot	Inh	Inh
[**Type**]	Ftv	Null	Null

5.3.3 Eventuality AOs

When **eventuality AOs** occur with attribution, it conveys an agent's intention/attitude towards a considered event, state or action. Eventuality AOs occur with "control" types of attribution ("Ctrl" for short), conveyed by any of three different classes of control verbs (Sag and Pollard, 1991). The first kind is anchored by a *verb of influence* like *persuade*, *permit*, *order*, and involve one agent influencing another agent to perform (or not perform) an action. The second kind is anchored by a *verb of commitment* like *promise*, *agree*, *try*, *intend*, *refuse*, *decline*, and involve an agent committing to perform (or not perform) an action. The third kind is anchored by a *verb of orientation* like *want*, *expect*, *wish*, *yearn*, and involve desire, expectation, or some similar mental orientation towards some state(s) of affairs. These sub-distinctions are not encoded in the annotation, but we have used the definitions as a guide for identifying these predicates. An example of the control attribution relation anchored by a verb of influence is given in (161).[25]

(161) Eward and Whittington had planned to leave the bank earlier, but Mr. Craven had persuaded them *to remain* until **the bank was in a healthy position**. (1949)

[25] While our use of the term *source* applies literally to agents responsible for the truth of a proposition, we continue to use the same term for the agents for facts and eventualities. Thus, for facts, the *source* represents the bearers of attitudes/knowledge, and for considered eventualities, the *source* represents the bearer of intentions/attitudes.

	REL	Arg1	Arg2
[**Source**]	Ot	Inh	Inh
[**Type**]	Ctrl	Null	Null

5.4 Scopal polarity

The *scopal polarity* feature is annotated on relations and their arguments to identify cases where verbs of attribution are negated on the surface - syntactically (e.g., *didn't say*, *don't think*) or lexically (e.g., *denied*), but where the negation in fact **reverses** the polarity of the attributed relation or argument content (Horn, 1978). Example (162) illustrates such a case. The *but* clause entails an interpretation such as *I think it's not a main consideration*, for which the negation must take *narrow* scope over the embedded clause rather than the higher clause. In particular, the interpretation of the contrastive relation denoted by *but* requires that `Arg2` should be interpreted under the scope of negation.

(162) "*Having the dividend increases is a supportive element in the market outlook*, but I don't think **it's a main consideration**," he says. (0090)

	REL	Arg1	Arg2
[**Source**]	Ot	Inh	Ot
[**Type**]	Comm	Null	PAtt
[**Polarity**]	Null	Null	Neg

To capture such entailments with surface negations on attribution verbs, an argument of a connective is marked "Neg" for *scopal polarity* when the interpretation of the connective requires the surface negation to take semantic scope over the lower argument. Thus, in Example (162), *scopal polarity* is marked as "Neg" for `Arg2`. When the neg-lowered interpretations are not present, *scopal polarity* is marked as the default "Null" (such as for the relation and `Arg1` of Example 162).

Note that this surface negation can be interpreted as taking scope only over the relation, rather than any argument as well. Since we have not observed this in the PDTB, we describe this case with the constructed example in (163). What the example shows is that in addition to entailing (163b) – in which case it would be annotated parallel to Example (162) above – (163a) can also entail (163c), such that the negation is intrepreted as taking semantic scope over the relation (Lasnik, 1975), rather than one of the arguments. As the *scopal polarity* annotations for (163c) show, lowering of the surface negation to the relation is marked as "Neg" for the *scopal polarity* of the relation.

(163) a. John doesn't think *Mary will get cured* because **she took the medication**.

b. John thinks *that* because **Mary took the medication**, *she will not get cured.*

	REL	Arg1	Arg2
[**Source**]	Ot	Inh	Inh
[**Type**]	PAtt	Null	Null
[**Polarity**]	Null	Neg	Null

c. John thinks *that Mary will get cured* not because **she took the medication** (but because she has started practising yoga.)

	REL	Arg1	Arg2
[**Source**]	Ot	Inh	Inh
[**Type**]	PAtt	Null	Null
[**Polarity**]	Neg	Null	Null

We note that *scopal polarity* does not capture the appearance of (opaque) internal negation that may appear on arguments or relations themselves. For example, a modified connective such as *not because* does not take "Neg" as the value for *scopal polarity*, but rather "Null". This is consistent with our goal of marking *scopal polarity* only for lowered negation, i.e., when surface negation from the attribution is lowered to either the relation or argument for interpretation.

5.5 Determinacy

The *determinacy* feature captures the fact that the attribution over a relation or argument can **itself** be cancelled in particular contexts, such as within negated, conditional, and infinitive contexts. Such indeterminacy is indicated by the value "Indet", while determinate contexts are simply marked by the default "Null". The annotation in Example (164) illustrates a case of indeterminacy of the (belief) attribution on the relation. Here, it is not that a belief or opinion about *our teachers educating our children better if only they got a few thousand dollars a year more* is being attributed to anyone, even "Arb" (ie, an arbitrary individual). Rather, the attribution is only being conjectured as a possibility. This indeterminacy is created by the infinitival context in which the attribution is embedded.

(164) It is silly libel on our teachers to think *they would educate our children better* if only **they got a few thousand dollars a year more**. (1286)

	REL	Arg1	Arg2
[**Source**]	Arb	Inh	Inh
[**Type**]	PAtt	Null	Null
[**Polarity**]	Null	Null	Null
[**Determinacy**]	Indet	Null	Null

5.6 Attribution spans

In addition to annotating the properties of attribution in terms of the features discussed above, we also annotate the *text span* associated with the attribution. The text span is annotated as a single (possibly discontinuous) complex reflecting the annotated features, and also includes all non-clausal modifiers of the elements contained in the span, for example, adverbs and appositive NPs. Connectives, however, may be excluded from the span. Example (165) shows a discontinuous annotation of the attribution, where the parenthetical *he argues* is excluded from the attribution phrase *the other side knows*, corresponding to the factive attribution.

(165) The other side, he argues, knows *Giuliani has always been pro-choice*, even though **he has personal reservations**. (0041)

	REL	Arg1	Arg2
[**Source**]	Ot	Inh	Inh
[**Type**]	Ftv	Null	Null
[**Polarity**]	Null	Null	Null
[**Determinacy**]	Null	Null	Null

We note that in annotating the attribution span as a single complex, we assume that the text anchors of the individual elements of the attribution - the *source*, *type*, *scopal polarity* and *determinacy* - can be identified by independent means with the help of other resources, such as the semantic role annotations (namely, Propbank (Kingsbury and Palmer, 2002)) on the Penn Treebank.

Spans for implicit writer attributions are left unmarked since there is no corresponding text that can be selected. The absence of a span annotation is simply taken to reflect writer attribution, together with the "Wr" value on the source feature.

Recognizing attributions is not trivial since they are often left unexpressed in the sentence in which the AO is realized, and have to be inferred from the prior discourse. For example, in (166), the relation and its arguments in the third sentence are attributed to Larry Shapiro, but this attribution is implicit and must be inferred from the first sentence. The spans for such implicit "Ot" attributions mark the text that provides the inference of the implicit attribution, which is just the closest occurrence of the explicit attribution phrase in the prior text.

(166) "There are certain cult wines that can command these higher prices,"
says Larry Shapiro of Marty's, ... "What's different is that it is happening with young wines just coming out. *We're seeing it* partly because **older vintages are growing more scarce**." (0071)

	REL	Arg1	Arg2
[**Source**]	Ot	Inh	Inh
[**Type**]	Comm	Null	Null
[**Polarity**]	Null	Null	Null
[**Determinacy**]	Null	Null	Null

The final aspect of the span annotation is that we also annotate non-clausal phrases as the anchors of attribution, such as prepositional phrases like *according to X*, and adverbs like *reportedly*, *allegedly*, *supposedly*. One such example is shown in (167). Note that while a specific individual is identified as the source of `Arg1` in this example, with "Ot" as the source value, many such phrases, especially the adverbs, refer to a non-specific generic source. In the latter case, the source value is marked as "Arb". Also, the type and scopal polarity of the attribution indicated by such phrasal attributions are assumed to be provided by the phrase itself. In (167), the *according to* preposition head of the attribution phrase is taken to reflect an assertion by the indicated agent, and the type is thus marked as "Comm".

(167) *No foreign companies bid on the Hiroshima project,* *according to the bureau*. But **the Japanese practice of deep discounting often is cited by Americans as a classic barrier to entry in Japan's market**. (0501)

	REL	Arg1	Arg2
[**Source**]	Wr	Ot	Inh
[**Type**]	Comm	Comm	Null
[**Polarity**]	Null	Null	Null
[**Determinacy**]	Null	Null	Null

For phrasal attributions, since the PDTB argument annotation guidelines do not allow for non-clausal modifiers of an argument to be excluded from the selection – a convention – they also appear as part of the argument span they modify. This is a slightly awkward aspect of the annotation, but since we also annotate attribution spans, it should be straightforward, if necessary, to strip away phrasal attribution spans when they appear *contained* within argument spans.

6 Description of PDTB representation format

6.1 Introduction

This section describes the representation format of the annotations of PDTB-2.0. and how they are linked to the Wall Street Journal corpus and the Penn Treebank annotations. Section 6.2 describes the default directory structure, and the mechanisms used to link the files. Section 6.3 describes the format of the PDTB annotation files. Section 6.4 gives an overview of the procedure used to link the PDTB annotations to the PTB. Relevant APIs, browsers and additional tools for viewing and querying the corpus are available from `http://www.seas.upenn.edu/~pdtb`.

6.2 Directory structure and linking mechanism

1. RAW refers to the Wall Street Journal raw text. RAW is assumed to have 25 sections, each with at most 100 files. Within the directory RawRoot, the naming conventions are such that RawRoot/00 is the RAW sub-directory for section 00, and RawRoot/00/wsj_0003 is the RAW file for section 00, file 03.

2. PTB refers to the Penn Treebank. PTB files are assumed to be in symbolic expression form, and PtbRoot/00/wsj_0003.mrg contains the parse trees for RawRoot/00/wsj_0003.

3. PDTB refers to the Penn Discourse Treebank. Within the directory PdtbRoot, the naming conventions are such that PdtbRoot/00 is the PDTB sub-directory for RawRoot section 00, and PdtbRoot/00/wsj_0003.pdtb contains the PDTB annotations for RawRoot/00/wsj_0003. In the LDC PDTB distribution, PdtbRoot is located within the top level *data* directory.

4. Given a PDTB file PdtbRoot/ij/wsj_ijkl.pdtb, the associated RAW file is RawRoot/ij/wsj_ijkl and the associated PTB file is PtbRoot/ij/wsj_ijkl.mrg.

5. PDTB files are linked to RAW files using *spans*. A span $p..q$ denotes the string in the associated RAW file starting from character p (inclusive) to character q (exclusive). For example, given the string *hello*, the span 0..1 is *h*, and the span 1..3 is the string *el*. A *span list* has the form $p_1..q_1;p_2..q_2...;p_n..q_n$, where $q_i \leq p_{i+1}$. Given the string *hello*, the span list 1..3;4..5 denotes *el o*. (Note the space added between spans.)

6. PDTB files are linked to PTB files using *Gorn addresses*. A Gorn address a_1, a_2, ... a_{n-1}, a_n denotes the a_nth child of the a_{n-1}th child of ... the a_2*th* child of the sentence number a_1 in the associated PTB file, and $T(a_1, a_2, ... a_n)$ denotes the subtree rooted at a_n.

   For example, given a PTB file with two sentences

   ```
   ((S0 (A a) (B b)))
   ((S1 (C c) (D d)))
   ```

   Gorn address 0,0 refers to the node *A*, and Gorn address 1,1,0 refers to the node *d*. Corresponding to this, $T(0,0)$ denotes the subtree rooted at *A*, i.e. (Aa), $T(0,1)$ denotes the subtree

rooted at B, i.e. (Bb), and T(1,1,0) denotes the subtree rooted at terminal d. In this presentation, we only refer to subtrees rooted at a particular node, and never to a node in isolation. That is, we simply write 0,0 to denote the subtree rooted at A, and we use the phrase *referring to a node* to mean *referring to the subtree rooted at the node*. Note that a label is a property of the node: There is no separate address for a node label.

A single number such as 0 refers to (the subtree rooted at) $S0$, and 1 refers to (the subtree rooted at) $S1$. Let G_1, G_2, ... G_n be Gorn addresses, then a *Gorn address list* is given by $G_1;G_2...;G_n$. A Gorn address list of length n denotes n nodes/subtrees. In such a list, we assume that G_i is not a prefix of G_j for all $1 \leq i,j \leq n$. (That is, G_i does not denote an ancestor of G_j.)

6.3 File format

Each PDTB file contains a list of relations. The following is the BNF description of its format. Non-terminals start with lower case, terminals start with upper case and ϵ denotes the empty production. (Some of the terminals here are types rather than specific tokens. Which are which will become clear in the relevant subsections below.)

```
 relationList ::= relation relationList | relation

relation ::=
    Explicit explicitRelation
  | Implicit implicitRelation
  | AltLex altLexRelation
  | EntRel entityRelation
  | NoRel noRelation

explicitRelation ::= selection explicitRelationFeatures sup arg arg sup

altLexRelation ::= selection altLexRelationFeatures sup arg arg sup

implicitRelation ::= inferenceSite implicitRelationFeatures sup arg arg sup

entityRelation ::= inferenceSite arg arg

noRelation ::= inferenceSite arg arg

sup ::= Sup selection | ϵ

arg ::= Arg selection attributionFeatures

selection ::= SpanList GornAddressList RawText
```

```
inferenceSite ::= StringPosition SentenceNumber

explicitRelationFeatures ::=
    attributionFeatures ConnHead semanticClass

altLexRelationFeatures ::=
    attributionFeatures semanticClass

implicitRelationFeatures ::=
    attributionFeatures Conn1 semanticClass Conn2 semanticClass
  | attributionFeatures Conn1 semanticClass

semanticClass ::= SemanticClass1 | SemanticClass1 SemanticClass2

attributionFeatures ::=
    Source Type Polarity Determinacy selection
  | Source Type Polarity Determinacy
```

6.3.1 General outline

The general outline of the data format for `Explicit`, `AltLex` and `Implicit` relations is shown below. While the format for the data in Sup1, Arg1, Arg2 and Sup2 is the same for all three types of relation, the relation-level data that appears under the header `____TYPE-OF-RELATION___`) differs, as will be explained below. (N.B. Each subcategory – TYPE-OF-RELATION, Sup1, Arg1, Arg2, Sup2 – begins with a similar, underscore-enclosed header.)

```
------------------------------------------------------------
____TYPE-OF-RELATION____ (ie. Explicit, AltLex, Implicit)
     relation data
____Sup1____
     sup1 data
____Arg1____
     arg1 data (including attribution)
____Arg2____
     arg2 data (including attribution)
____Sup2____
     sup2 data
------------------------------------------------------------
```

For `EntRel` and Norel, the general outline is:

```
------------------------------------------------------------
____TYPE-OF-RELATION____ (ie. Entrel, NoRel)
```

```
     relation data
____Arg1____
     arg1 data (NO attribution)
____Arg2____
     arg2 data (NO attribution)
------------------------------------------------------
```

`EntRel` and `NoRel` contain neither Sup1 and Sup2, nor attribution features for Arg1 and Arg2 (cf. Section 6.3.5).

6.3.2 `Explicit` relation

The following is the outline of the file format for an `Explicit` relation. Items denoted in *italics* are optional, and their values are only printed if they exist in the annotations.

```
------------------------------------------------------
____Explicit____
SpanList
GornAddress
#### Text ####
RawText
###############
#### Features ####
Source, Type, Polarity, Determinacy
SpanList (for relation attribution if available)
GornAddress (for relation attribution if available)
#### Text ####
RawText (for relation attribution if available)
###############
ConHead, SemanticClass1, SemanticClass2
___Sup1___
SpanList
GornAddress
#### Text ####
RawText
###############
____Arg1____
SpanList
GornAddress
#### Text ####
RawText
###############
#### Features ####
Source, Type, Polarity, Determinacy
```

SpanList (for Arg1 attribution if available)
GornAddress (for Arg1 attribution if available)
Text
RawText (for Arg1 attribution if available)
##############
____Arg2____
SpanList
GornAddress
Text
RawText
##############
Features
Source, Type, Polarity, Determinacy
SpanList (for Arg2 Attribution if available)
GornAddress (for Arg2 Attribution if available)
Text
RawText (for Arg2 Attribution if available)
##############
___Sup2___
SpanList
GornAddress
Text
RawText
##############

Comments:

1. SpanList and GornAddressList were explained in Section 6.2. The SpanList corresponds to selections made by the annotator, while the GornAddressList is computed programmatically given the SpanList, as described in Section 6.4.

2. A pair of SpanList and GornAddressList values will always be associated with a RawText, enclosed in the following structure:

   ```
   ### Text ####
   RawText
   #############
   ```

   RawText refers to the portion of text from the Wall Street Journal file corresponding to the SpanList selection made by the annotator. Note that the portion of text will reflect the formatting of the original Wall Street Journal file. If the text is contains line breaks, then RawText will reflect that as well. Usually, most RawText are contained wtihin a single line.

3. The subheader `#### Features ####` denotes a number of relevant values, depending on where it appears:

   Following the relation-level header `____Explicit____` relevant values include:

   - attribution values (Source, Type, Polarity, Determinacy) and if relevant, attribution Span-list, GornAddress and RawText;
   - ConnHead and Semantic classes (see below).

   Following the argument headers `____Arg1____` and `____Arg2____` are features denoting attribution values only.

4. ConnHead is the head of the `Explicit` connective. In most cases, ConnHead is equivalent to the actual RawText of the `Explicit` connective. In the case of modified connectives such as *largely because*, ConnHead has the value *because* while the RawText includes the entire modifier+connective complex *largely because*.

5. A ConnHead is obligatorily associated with a sense (SemanticClass1). Optionally, it may have a second sense (SemanticClass2). Sense labels reflect the full hierarchical classification of the sense, shown in the order "Class.Type.SubType". (See Section 4 for an explanation of the hierarchical classification.) For example, the label "Contingency.Cause.Reason" refers to the *Reason* subtype of the *Cause* type of the *Contingency* class.

A sample `Explicit` relation from the corpus is shown below. Of particular interest is that:

- ConnHead is *because* and RawText is *largely because*.
- Neither Sup1 nor Sup2 (optional material) is present in the relation.
- There are selection features (SpanList, GornAddresses and RawText) for the *attribution* of the relation, but not for the arguments, since their *attribution* is inherited from that of the relation.

```
------------------------------------------------------
____Explicit____
2084..2099
13,1,1,1,1,1,1,2,0;13,1,1,1,1,1,1,2,1
#### Text ####
largely because
###############
#### Features ####
Ot, Comm, Null, Null
2007..2038
13,0;13,1,0;13,2
#### Text ####
Mr. Lothson of PaineWebber said
###############
because, Contingency.Cause.Reason
```

```
____Arg1____
2039..2083
13,1,1,0;13,1,1,1,0;13,1,1,1,1,0;13,1,1,1,1,1,0;13,1,1,1,1,1,1,0;13,1,1,1,1,1,1,1
#### Text ####
the company's sales pace has been picking up
##############
#### Features ####
Inh, Null, Null, Null
____Arg2____
2100..2193
13,1,1,1,1,1,1,2,2
#### Text ####
the effect of unfavorable exchange rates has been easing -- a pattern continuing
this quarter
##############
#### Features ####
Inh, Null, Null, Null
-----------------------------------------------------
```

6.3.3 `AltLex` relation

The following is the outline of the file format for an `AltLex` relation:

```
-----------------------------------------------------
____AltLex____
SpanList
GornAddress
#### Text ####
RawText
##############
#### Features ####
Source, Type, Polarity, Determinacy
SpanList (for relation attribution if available)
GornAddress (for relation attribution if available)
#### Text ####
RawText (for relation attribution if available)
##############
SemanticClass1, SemanticClass2
___Sup1___
SpanList
GornAddress
#### Text ####
RawText
##############
```

____Arg1____
SpanList
GornAddress
Text
RawText
##############
Features
Source, Type, Polarity, Determinacy
SpanList (for Arg1 attribution if available)
GornAddress (for Arg1 attribution if available)
Text
RawText (for Arg1 attribution if available)
##############
____Arg2____
SpanList
GornAddress
Text
RawText
##############
Features
Source, Type, Polarity, Determinacy
SpanList (for Arg2 Attribution if available)
GornAddress (for Arg2 Attribution if available)
Text
RawText (for Arg2 Attribution if available)
##############
____Sup2____
SpanList
GornAddress
Text
RawText
##############

Comments:

1. Note that an `AltLex` relation differs from an `Explicit` relation in simply not having a ConnHead category. The `AltLex` relation only has an obligatory SemanticClass1 and an optional SemanticClass2.

A sample `AltLex` relation from the corpus is shown below. In this example, note the selection features (SpanList, GornAddress and RawText) for the attribution of Arg1.

```
____AltLex____
3487..3499
32,0;32,1,0
#### Text ####
Both reflect
##############
#### Features ####
Wr, Comm, Null, Null
Contingency.Cause.Reason
____Arg1____
3365..3485
31,1,1
#### Text ####
the average pay of its clients fell to $66,743 last year from $70,765 in 1987;
severance pay dropped to 25 weeks from 29
##############
#### Features ####
Ot, Comm, Null, Null
3319..3364
31,0;31,1,0;31,2
#### Text ####
Outplacement consultant Right Associates says
##############
____Arg2____
3487..3558
32
#### Text ####
Both reflect the dismissal of lower-level and shorter-tenure executives
##############
#### Features ####
Inh, Null, Null, Null
-------------------------------------------------------
```

6.3.4 `Implicit` relation

The following is the outline of the file format for an `Implicit` relation. Once again, optional material is in *italics*:

```
-------------------------------------------------------
____Implicit____
StringPosition
SentenceNo
#### Features ####
Source, Type, Polarity, Determinacy
```

SpanList (for relation attribution if available)
GornAddress (for relation attribution if available)
Text
RawText (for relation attribution if available)
###############
Conn1, SemanticClass1, *SemanticClass2*
Conn2, SemanticClass1, SemanticClass2
____Sup1____
SpanList
GornAddress
Text
RawText
###############
_____Arg1_____
SpanList
GornAddress
Text
RawText
###############
Features
Source, Type, Polarity, Determinacy
SpanList (for Arg1 attribution if available)
GornAddress (for Arg1 attribution if available)
Text
RawText (for Arg1 attribution if available)
###############
_____Arg2_____
SpanList
GornAddress
Text
RawText
###############
Features
Source, Type, Polarity, Determinacy
SpanList (for Arg2 Attribution if available)
GornAddress (for Arg2 Attribution if available)
Text
RawText (for Arg2 Attribution if available)
###############
____Sup2____
SpanList
GornAddress
Text

RawText
##############

Comments:

1. StringPosition and SentenceNo give the site of inference of the `Implicit` connective. The StringPosition is the offset of the first character of Arg2 of the `Implicit` connective and SentenceNo is the sentence number of Arg2.

2. An annotator may infer at most two `Implicit` connectives for an implicit relation. The obligatory first connective is Conn1. It must have at least one sense (SemanticClass1) and optionally a second sense SemanticClass2. If the annotator optionally chose a second `Implicit` connective for the relation, this connective is recorded as Conn2. If a Conn2 is present, it must at least have one sense (SemanticClass1) and optionally a second sense (SemanticClass2).

A sample `Implicit` relation from the corpus is shown below. Note that there is a Conn1 with an obligatory SemanticClass1, but it does not have the optional SemanticClass2. The optional Conn2 along with its corresponding SemanticClass1 and SemanticClass2 are absent.

```
---------------------------------------------------------
____Implicit____
419
4
#### Features ####
Wr, Comm, Null, Null
for example, Expansion.Instantiation
____Arg1____
281..306
3
#### Text ####
Others were more cautious
##############
#### Features ####
Inh, Null, Null, Null
____Arg2____
419..486
4,4,1
#### Text ####
that eroding confidence might undermine future economic development
##############
#### Features ####
Ot, Comm, Null, Null
308..418
```

```
4,0;4,1;4,2;4,3;4,4,0;4,5
#### Text ####
In a July analysis titled ‘‘From Euphoria to Despair,’’ W.I. Carr (Far East) Ltd.,
another securities firm, said
##############
------------------------------------------------------
```

6.3.5 `EntRel` and `NoRel`

The file format for `EntRel` or `NoRel` do not contain any Sups or Features. In other respects, it is similar to the structure for `Implicit` relations.

```
------------------------------------------------------
____EntRel____ (or NoRel)
StringPosition
SentenceNo
____Arg1____
SpanList
GornAddress
#### Text ####
RawText
##############
____Arg2____
SpanList
GornAddress
#### Text ####
RawText
##############
------------------------------------------------------
```

A sample `EntRel` from the corpus is shown below. Note the absence of Sups and features.

```
------------------------------------------------------
____EntRel____
7481
55
____Arg1____
7419..7479
54
#### Text ####
Do they want the spotlight for themselves or for their cause
##############
____Arg2____
```

```
7481..7634
55
#### Text ####
Here is a good rule of thumb: If the movement produced the leader, the chance that
he is sincere is much greater than if the leader produced the movement
##############
-------------------------------------------------------
```

6.4 Computation of Gorn addresses

As noted earlier, a GornAddressList (of PTB nodes) is computed programmatically from the SpanList selections made by the annotator. Here we briefly describe that computation. In this description, Γ is used to denote a PTB node whose yield consists of only punctuation or traces, while γ is used to denote a clausal PTB node – one whose label starts with *S*, *PRN* (if it has a child whose label starts with *S*), or *PP* (if it has a child whose labels starts with *S*). The arguments selected in the PDTB correspond in most cases to γ nodes. However, when annotators select a span in the RAW files, one needs to consider which γ nodes on its periphery need to be included. For example, consider the following (artificial) scenario:

```
RAW:

John goes home early, when he is tired.

PTB:

((S
    (NP (NN John))
    (VP (V goes)
        (NP (NN home))
        (ADVP (RB early))
        (, ,)
        (SBAR (WHADVP_1 (RB when)
                        (S
                            (NP (PRP he))
                            (VP (V is)
                                (ADJP (JJ tired))
                                (ADVP-TMP (-NONE- *T*-1))
                            )
                        )
              )
        )
    )
```

```
    (. .)
))

PDTB:

Conn: 22..26 (when)
Arg1: 0..20 (John goes home early)
Arg2: 27..38 (he is tired)
```

In this case, if the minimal set of nodes corresponding to the RAW selections were computed, the γ nodes corresponding to the comma, ADVP trace and period would be excluded. These cases are handled in two steps:

1. *Span stretching* - The SpanList selections are stretched (in both directions) to include Γ nodes on their periphery, as long as (i) no other lexical item is included, and (ii) all γ nodes are respected. The latter constraint ensures that no additional γ nodes are included (clausal nodes of the type *S*, *PRN* or *PP*) that were not contained within the original annotator selection.

   In the example, the constraints ensure that the Γ node corresponding to the period is not included in Arg2, because to include this, the *SBAR* node (which is a γ node) would have to be included, and this is not contained within the annotator's selection.

   The GornAddressList, at the end of this phase, denotes the highest set of nodes that dominate the stretched span exactly. For Conn, this would be 0,1,4,0,0. For Arg1: 0,0;0,1,0;0,1,1;0,1,2;0,1,3. For Arg2: 0,1,4,0,1. Note that the period with address 0,2 does (yet) not appear in any selection.

2. *Sibling inclusion* - For each node in Arg1, each Γ node sibling of that node that does not occur in Conn, Arg2, Sup1 or Sup2 should be added to Arg1. Similarly for Arg2, Sup1 and Sup2 (in that order). This step results in the period with address 0,2 being added to Arg1.

Note that span stretcing and sibling inclusion are also carried out in the Gorn address computation of the attribution spans.

6.5 Acknowledgements

We would like to thank the following people:

1. **Annotators:** Lukasz Abramowicz, Dan Afergan, Soobia Afroz, Driya Amandita, John Bell, Alex Channer, Sara Clopton, George Cooper, Cassandre Creswell, Sarah Johnstone, John Laury, Alan Lee, Marielle Lerner, Sophia Malamud, Chris Moulton, Viraj Narayanan, Emily Pawley, Steven Pettington, Sami Saba, Adi Shifir, Sandhya Sundaresan, Nianwen Xue.
2. **Programmers:** Jeremy Lacivita, Tom Morton, Gerry Campion.
3. **Undergraduate and Graduate researchers:** John Bell, Mridhula Raghupathy and Hena Mehta.

Appendix A

This Appendix provides a distribution and counts of the types of `Explicit` connectives in PDTB-2.0., along with their sense types. The full distribution is split across Tables 2–7. There are 100 distinct types of `Explicit` connectives (but see a discussion of multiple connectives in Section 1.5), given in the first column. The total number of `Explicit` connective tokens annotated is 18459 (the Total for the second as well as the third columns across all the tables). Counts are given for the type of connective (third column) as well as for each of its labeled senses (in parentheses in the second column). Multiple senses for a connective are shown as a distinct type of sense, the different senses separated by "/". See Figure 1 for the hierarchical classification of the senses shown here. Note that modified forms of `Explicit` connectives are treated as the same type as that of the head and are therefore not shown separately in these tables. The association between senses and the full forms of connectives can be found in Appendix B.

`Explicit` **Connective**	**Senses**	**Total**
accordingly	result (5)	5
additionally	Conjunction (7)	7
after	expectation (2), expectation/succession (1), reason/succession (50), specification/succession (1), succession (523)	577
afterward	precedence (11)	11
also	Conjunction (1733), Conjunction/Synchrony (2), List (10), specification (1)	1746
alternatively	Alternative (2), disjunctive (4)	6
although	COMPARISON (16), Concession (1), contra-expectation (21), Contrast (114), Exception (1), expectation (132), juxtaposition (34), opposition (9)	328
and	Conjunction (2543), Conjunction/Contrast (1), Conjunction/general (1), Conjunction/juxtaposition (5), Conjunction/opposition (1), Conjunction/precedence (30), Conjunction/result (138), Conjunction/Synchrony (4), contra-expectation (1), Contrast (3), EXPANSION (1), generalization (1), hypothetical (1), hypothetical/precedence (1), Instantiation (1), juxtaposition (11), juxtaposition/List (1), List (210), opposition (5), precedence (1), precedence/result (1), result (38), specification (1)	3000
as	COMPARISON/Synchrony (2), Conjunction (1), CONTINGENCY/Synchrony (1), contra-expectation (1), EXPANSION (1), expectation (2), justification (1), justification/Synchrony (5), juxtaposition/Synchrony (2), opposition (1), reason (166), reason/succession (23), reason/Synchrony (144), reason/TEMPORAL (1), result/Synchrony (2), succession (3), Synchrony (387)	743
as a result	result (78)	78

Table 2: `Explicit` Connectives and their senses *(Cont. on next page)*

Explicit **Connective**	**Senses**	**Total**
as an alternative	disjunctive (2)	2
as if	contra-expectation (4), EXPANSION (11), Pragmatic_Concession (1)	16
as long as	factual_present/Synchrony (4), general (6), general/Synchrony (1), hypothetical (7), hypothetical/Synchrony (2), Synchrony (4)	24
as soon as	succession (11), Synchrony (9)	20
as though	COMPARISON (3), specification (2)	5
as well	Conjunction (6)	6
because	implicit_assertion (2), justification (2), reason (854)	858
before	precedence (323), succession (3)	326
before and after	Asynchronous (1)	1
besides	Conjunction (17), conjunctive (1), juxtaposition (1)	19
but	chosen_alternative (2), COMPARISON (260), COMPARISON/Conjunction (2), COMPARISON/precedence (1), Concession (2), Conjunction (63), Conjunction/contra-expectation (1), Conjunction/Contrast (1), Conjunction/juxtaposition (1), Conjunction/Pragmatic_Concession (4), Conjunction/Pragmatic_Contrast (14), contra-expectation (494), contra-expectation/juxtaposition (1), Contrast (1609), Contrast/precedence (2), Exception (2), EXPANSION (2), EXPANSION/Pragmatic_Contrast (1), expectation (12), juxtaposition (636), juxtaposition/List (1), opposition (174), opposition/precedence (1), opposition/reason (1), Pragmatic_Concession (4), Pragmatic_Contrast (14), Pragmatic_Contrast/specification (1), result (1), specification (1)	3308
by comparison	COMPARISON (1), Contrast (3), juxtaposition (7)	11
by contrast	Contrast (11), juxtaposition (12), opposition (4)	27
by then	precedence (1), precedence/reason (1), succession (3), Synchrony (2)	7
consequently	result (10)	10
conversely	Contrast (2)	2
earlier	juxtaposition/succession (1), succession (14)	15
either..or	Alternative (1), conjunctive (1), disjunctive (2)	4
else	Alternative (1)	1
except	disjunctive (1), Exception (9)	10
finally	Conjunction (11), Conjunction/precedence (1), EXPANSION/precedence (1), List (7), precedence (10), precedence/result (2)	32
for	reason (3)	3
for example	Instantiation (194), specification (2)	196
for instance	Instantiation (98)	98
further	Conjunction (9)	9
furthermore	Conjunction (11)	11
hence	result (4)	4

Table 3: Explicit Connectives and their senses *(Cont. on next page)*

Explicit **Connective**	**Senses**	**Total**
however	COMPARISON (49), Concession (5), Conjunction (2), contra-expectation (70), Contrast (234), expectation (2), juxtaposition (89), List/opposition (1), opposition (31), Pragmatic_Concession (1), Pragmatic_Contrast (1)	485
if	COMPARISON (1), Concession (1), Condition (2), contra-expectation (3), contra-expectation/general (1), contra-expectation/hypothetical (3), Contrast (1), expectation (34), expectation/factual_present (2), expectation/general (1), expectation/hypothetical (1), factual_past (9), factual_present (73), factual_present/juxtaposition (1), general (175), general/Synchrony (1), hypothetical (682), hypothetical/specification (1), implicit_assertion (29), juxtaposition (4), Pragmatic_Contrast (1), relevance (20), result (1), specification (1), unreal_past (53), unreal_present (122)	1223
if and when	general (1), hypothetical (1), hypothetical/Synchrony (1)	3
if..then	factual_present (9), general (5), hypothetical (22), implicit_assertion (1), unreal_past (1)	38
in addition	Conjunction (165)	165
in contrast	Contrast (5), juxtaposition (5), opposition (2)	12
in fact	COMPARISON (1), Conjunction (33), Contrast (2), EXPANSION (2), generalization (1), Instantiation (2), juxtaposition (3), Restatement (2), specification (36)	82
in other words	equivalence (10), generalization (3), Restatement (4)	17
in particular	Instantiation (6), specification (9)	15
in short	generalization (4)	4
in sum	generalization (2)	2
in the end	Conjunction (1), contra-expectation (1), EXPANSION (1), generalization (2), juxtaposition (1), precedence (1), result (1), specification (1)	9
in turn	Asynchronous (1), Conjunction (4), Conjunction/opposition (1), Conjunction/precedence (1), Conjunction/Synchrony (1), precedence (14), precedence/result (4), result (3), specification (1)	30
indeed	COMPARISON (1), Conjunction (53), Conjunction/result (1), equivalence (1), EXPANSION (3), generalization (1), Instantiation (1), justification (4), Restatement (7), specification (32)	104
insofar as	reason (1)	1
instead	Alternative (1), Alternative/chosen_alternative (3), chosen_alternative (105), Contrast (2), juxtaposition (1)	112
later	Conjunction (1), precedence (90)	91
lest	disjunctive (1), hypothetical (1)	2
likewise	Conjunction (8)	8
meantime	Asynchronous (1), Conjunction/succession (1), Conjunction/Synchrony (1), juxtaposition/Synchrony (1), succession (1), Synchrony (10)	15

Table 4: Explicit Connectives and their senses *(Cont. on next page)*

Explicit **Connective**	**Senses**	**Total**
meanwhile	Conjunction (25), Conjunction/Contrast (2), Conjunction/juxtaposition (2), Conjunction/Synchrony (92), Conjunction/TEMPORAL (1), contra-expectation (1), Contrast/Synchrony (6), Contrast/TEMPORAL (1), juxtaposition (9), juxtaposition/List (2), juxtaposition/Synchrony (15), opposition (2), opposition/Synchrony (9), Synchrony (26)	193
moreover	Conjunction (100), List (1)	101
much as	chosen_alternative (1), chosen_alternative/Synchrony (1), COMPARISON (1), expectation (2), specification (1)	6
neither..nor	Conjunction (1), disjunctive (1), opposition (1)	3
nevertheless	COMPARISON (1), Conjunction/Pragmatic_Concession (1), contra-expectation (19), Contrast (11), juxtaposition (7), opposition (1), Pragmatic_Concession (4)	44
next	Conjunction/precedence (1), precedence (6)	7
nonetheless	COMPARISON (5), Conjunction (1), contra-expectation (17), Contrast (4)	27
nor	Alternative/Conjunction (1), Alternative/conjunctive (1), Conjunction (25), conjunctive (3), Contrast (1)	31
now that	precedence/reason (1), reason (9), reason/succession (4), reason/Synchrony (4), result (2), Synchrony (2)	22
on the contrary	Contrast (2), juxtaposition (1), opposition (1)	4
on the one hand..on the other hand	Contrast (1)	1
on the other hand	COMPARISON (1), Conjunction (1), Contrast (14), juxtaposition (12), opposition (9)	37
once	general (3), general/succession (1), hypothetical (1), hypothetical/succession (2), reason/succession (7), result/succession (1), succession (67), Synchrony (2)	84
or	Alternative (32), Alternative/conjunctive (2), Conjunction (5), conjunctive (41), Contrast (1), disjunctive (12), equivalence (1), EXPANSION (1), implicit_assertion (2), Restatement (1)	98
otherwise	conjunctive (1), disjunctive (21), Exception (2)	24
overall	Conjunction (3), generalization (2), Restatement (3), specification (4)	12
plus	Conjunction (1)	1
previously	juxtaposition/succession (3), succession (46)	49
rather	chosen_alternative (6), Contrast (3), juxtaposition (1), specification (7)	17
regardless	contra-expectation (2)	2
separately	Alternative/Conjunction (1), Conjunction (69), Conjunction/Contrast (1), Conjunction/Synchrony (2), List (1)	74
similarly	COMPARISON/Conjunction (2), Conjunction (16)	18
simultaneously	Synchrony (6)	6
since	precedence (1), reason (94), reason/succession (10), succession (78), Synchrony (1)	184

Table 5: Explicit Connectives and their senses *(Cont. on next page)*

Explicit Connective	Senses	Total
so	reason (1), result (262)	263
so that	result (31)	31
specifically	Conjunction (1), specification (9)	10
still	COMPARISON (8), Concession (1), contra-expectation (80), Contrast (68), expectation (2), juxtaposition (23), opposition (5), precedence (2), Synchrony (1)	190
then	Alternative (1), COMPARISON (1), Conjunction (10), Conjunction/precedence (1), Conjunction/result (1), contra-expectation (1), Contrast (1), List (2), List/precedence (1), precedence (302), precedence/reason (1), precedence/result (7), reason (1), result (4), Synchrony (5), unreal_present (1)	340
thereafter	opposition/precedence (1), precedence (9), succession (1)	11
thereby	result (12)	12
therefore	result (26)	26
though	COMPARISON (12), Concession (6), contra-expectation (45), Contrast (103), expectation (105), juxtaposition (46), opposition (3)	320
thus	result (112)	112
till	precedence (3)	3
ultimately	Conjunction (1), Conjunction/precedence (2), EXPANSION/precedence (1), EXPANSION/succession (1), precedence (9), precedence/Restatement (1), precedence/result (2), reason/Restatement (1)	18
unless	disjunctive (94), hypothetical (1)	95
until	disjunctive (3), general (2), hypothetical (16), hypothetical/succession (2), precedence (133), succession (5), Synchrony (1)	162
when	CONTINGENCY/succession (1), CONTINGENCY/Synchrony (2), CONTINGENCY/TEMPORAL (1), contra-expectation (1), Contrast (1), disjunctive (2), expectation (8), expectation/Synchrony (4), factual_present (2), factual_present/specification (1), factual_present/Synchrony (5), general (100), general/precedence (1), general/specification (1), general/succession (9), general/Synchrony (50), general/TEMPORAL (3), hypothetical (11), hypothetical/succession (2), hypothetical/Synchrony (10), implicit_assertion (11), implicit_assertion/Synchrony (1), justification/succession (1), juxtaposition/Synchrony (1), opposition (1), opposition/Synchrony (1), precedence (4), reason (6), reason/succession (65), reason/Synchrony (39), reason/TEMPORAL (1), relevance/succession (1), result/succession (1), result/Synchrony (2), specification/Synchrony (2), succession (157), Synchrony (477), TEMPORAL (3)	989

Table 6: Explicit Connectives and their senses *(Cont. on next page)*

`Explicit` **Connective**	**Senses**	**Total**
when and if	hypothetical/succession (1)	1
whereas	Contrast (1), juxtaposition (3), opposition (1)	5
while	COMPARISON (18), COMPARISON/Synchrony (4), Concession (1), Conjunction (39), Conjunction/Contrast (1), Conjunction/juxtaposition (5), Conjunction/Synchrony (21), Conjunction/TEMPORAL (1), contra-expectation (3), Contrast (120), Contrast/Synchrony (22), expectation (79), expectation/Synchrony (3), juxtaposition (182), juxtaposition/List (9), juxtaposition/Synchrony (26), List (3), List/opposition (1), opposition (78), opposition/Synchrony (11), Synchrony (154)	781
yet	COMPARISON (12), Conjunction (2), Conjunction/contra-expectation (2), contra-expectation (32), Contrast (26), juxtaposition (21), opposition (6)	101

Table 7: `Explicit` Connectives and their senses

Appendix B

This Appendix provides a distribution of all the distinct senses annotated for `Explicit` connectives (first column). 111 distinct senses are recorded for `Explicit` connectives (but see the discussion of multiple connectives in Section 1.5. For each sense, the second column provides a list of the `Explicit` connectives for which the sense was annotated, with counts given for each connective (in parentheses). Note that unlike Appendix A, connectives listed in these tables show modified forms of connectives separately so that the correspondence between a sense and the exact form of the connective can be more clearly observed. The full distribution is split across Tables 8-13. Multiple senses annotated for a connective appear as their own type, separated by "/". See Figure 1 for the hierarchical classification of the senses shown here. The Total of the counts in the second and third columns is 18459, which is the total number of annotated `Explicit` connective tokens.

Sense	`Explicit` **Connectives**	**Total**
Alternative	alternatively (2), either..or (1), else (1), instead (1), or (32), then (1)	38
Alternative/chosen_alternative	instead (3)	3
Alternative/Conjunction	nor (1), separately (1)	2
Alternative/conjunctive	nor (1), or (2)	3
Asynchronous	before and after (1), in the meantime (1), in turn (1)	3
chosen_alternative	as much as (1), but (2), instead (105), rather (6)	114
chosen_alternative/Synchrony	so much as (1)	1
COMPARISON	although (16), as though (3), but (260), by comparison (1), even if (1), even though (2), however (49), in fact (1), indeed (1), much as (1), nevertheless (1), nonetheless (5), on the other hand (1), still (8), then (1), though (10), while (18), yet (12)	391
COMPARISON/Conjunction	but (2), similarly (2)	4
COMPARISON/precedence	but (1)	1
COMPARISON/Synchrony	even as (2), while (4)	6
Concession	although (1), but (2), even if (1), even though (5), however (5), still (1), though (1), while (1)	17
Condition	if (2)	2
Conjunction	additionally (7), also (1733), and (2543), as well (6), besides (17), but (63), even then (1), finally (11), further (9), furthermore (11), however (2), in addition (165), in fact (33), in the end (1), in turn (4), indeed (53), just as (1), later (1), likewise (8), meanwhile (25), moreover (100), neither..nor (1), nonetheless (1), nor (25), on the other hand (1), or (5), overall (3), plus (1), separately (69), similarly (16), specifically (1), then (9), ultimately (1), while (39), yet (2)	4968
Conjunction/contra-expectation	but (1), yet (2)	3
Conjunction/Contrast	and (1), but (1), meanwhile (2), separately (1), while (1)	6
Conjunction/general	and (1)	1
Conjunction/juxtaposition	and (5), but (1), meanwhile (2), while (5)	13
Conjunction/opposition	and (1), in turn (1)	2

Table 8: Senses and their associated `Explicit` Connectives *(Cont. on next page)*

Sense	Explicit Connectives	Total
Conjunction/Pragmatic_Concession	but (4), nevertheless (1)	5
Conjunction/Pragmatic_Contrast	but (14)	14
Conjunction/precedence	and (30), finally (1), in turn (1), next (1), then (1), ultimately (2)	36
Conjunction/result	and (138), indeed (1), then (1)	140
Conjunction/succession	in the meantime (1)	1
Conjunction/Synchrony	also (2), and (4), in turn (1), meantime (1), meanwhile (92), separately (2), while (21)	123
Conjunction/TEMPORAL	meanwhile (1), while (1)	2
conjunctive	besides (1), either..or (1), nor (3), or (41), otherwise (1)	47
CONTINGENCY/succession	when (1)	1
CONTINGENCY/Synchrony	as (1), when (2)	3
CONTINGENCY/TEMPORAL	when (1)	1
contra-expectation	although (21), and (1), as if (4), but (494), even as (1), even if (3), even still (1), even then (1), even though (15), however (70), in the end (1), meanwhile (1), nevertheless (19), nonetheless (17), regardless (2), still (79), though (30), when (1), while (3), yet (32)	796
contra-expectation/general	even if (1)	1
contra-expectation/hypothetical	even if (3)	3
contra-expectation/juxtaposition	but (1)	1
Contrast	although (114), and (3), but (1609), by comparison (3), by contrast (11), conversely (2), even though (15), however (234), if (1), in contrast (5), in fact (2), instead (2), nevertheless (11), nonetheless (4), nor (1), on the contrary (2), on the one hand..on the other hand (1), on the other hand (14), or (1), rather (3), still (68), then (1), though (88), when (1), whereas (1), while (120), yet (26)	2343
Contrast/precedence	but (2)	2
Contrast/Synchrony	meanwhile (6), while (22)	28
Contrast/TEMPORAL	meanwhile (1)	1
disjunctive	alternatively (4), as an alternative (2), either..or (2), except (1), except when (1), just until (1), lest (1), neither..nor (1), or (12), otherwise (21), unless (94), until (2), when (1)	143
equivalence	in other words (10), indeed (1), or (1)	12
Exception	although (1), but (2), except (9), otherwise (2)	14
EXPANSION	and (1), as (1), as if (11), but (2), in fact (2), in the end (1), indeed (3), or (1)	22
EXPANSION/Pragmatic_Contrast	but (1)	1
EXPANSION/precedence	finally (1), ultimately (1)	2
EXPANSION/succession	ultimately (1)	1
expectation	although (132), as much as (1), but (12), even after (2), even as (2), even if (31), even though (52), even when (5), however (2), if (3), much as (1), still (2), though (53), when (3), while (79)	380

Table 9: Senses and their associated Explicit Connectives *(Cont. on next page)*

Sense	Explicit Connectives	Total
expectation/factual_present	if (2)	2
expectation/general	even if (1)	1
expectation/hypothetical	even if (1)	1
expectation/succession	even after (1)	1
expectation/Synchrony	even when (4), while (3)	7
factual_past	if (9)	9
factual_present	even if (4), if (69), if..then (9), when (2)	84
factual_present/juxtaposition	if (1)	1
factual_present/specification	especially when (1)	1
factual_present/Synchrony	as long as (3), only as long as (1), when (5)	9
general	as long as (6), at least when (1), especially if (1), even if (2), if (170), if and when (1), if..then (5), once (3), only if (1), only when (1), particularly if (1), until (2), when (98)	292
general/precedence	when (1)	1
general/specification	especially when (1)	1
general/succession	even when (1), once (1), when (8)	10
general/Synchrony	as long as (1), if (1), only when (2), when (48)	52
general/TEMPORAL	when (3)	3
generalization	and (1), in fact (1), in other words (3), in short (4), in sum (2), in the end (2), indeed (1), overall (2)	16
hypothetical	and (1), as long as (7), especially if (3), even if (30), even when (1), if (632), if and when (1), if only (2), if..then (22), lest (1), once (1), only if (12), only when (2), particularly if (2), typically, if (1), unless (1), until (16), when (8)	743
hypothetical/precedence	and (1)	1
hypothetical/specification	especially if (1)	1
hypothetical/succession	once (2), until (2), when (2), when and if (1)	7
hypothetical/Synchrony	as long as (2), if and when (1), only when (1), when (9)	13
implicit_assertion	if (29), if..then (1), just because (2), or (2), when (11)	45
implicit_assertion/Synchrony	when (1)	1
Instantiation	and (1), for example (194), for instance (98), in fact (2), in particular (6), indeed (1)	302
justification	as (1), because (2), indeed (4)	7
justification/succession	when (1)	1
justification/Synchrony	as (5)	5
juxtaposition	although (34), and (11), besides (1), but (636), by comparison (7), by contrast (12), even though (5), however (89), if (4), in contrast (5), in fact (3), in the end (1), instead (1), meanwhile (9), nevertheless (7), on the contrary (1), on the other hand (12), rather (1), still (23), though (41), whereas (3), while (182), yet (21)	1109
juxtaposition/List	and (1), but (1), meanwhile (2), while (9)	13
juxtaposition/succession	earlier (1), previously (3)	4

Table 10: Senses and their associated Explicit Connectives *(Cont. on next page)*

Sense	Explicit Connectives	Total
juxtaposition/Synchrony	as (2), in the meantime (1), meanwhile (15), when (1), while (26)	45
List	also (10), and (210), finally (7), moreover (1), separately (1), then (2), while (3)	234
List/opposition	however (1), while (1)	2
List/precedence	then (1)	1
opposition	although (9), and (5), but (174), by contrast (4), even as (1), even though (1), however (31), in contrast (2), meanwhile (2), neither..nor (1), nevertheless (1), on the contrary (1), on the other hand (9), still (5), though (2), when (1), whereas (1), while (78), yet (6)	334
opposition/precedence	but (1), thereafter (1)	2
opposition/reason	but (1)	1
opposition/Synchrony	meanwhile (9), when (1), while (11)	21
Pragmatic_Concession	as if (1), but (4), however (1), nevertheless (4)	10
Pragmatic_Contrast	but (14), however (1), if only (1)	16
Pragmatic_Contrast/specification	but (1)	1
precedence	a day or two before (1), a decade before (1), a full five minutes before (1), a week before (1), about six months before (1), afterward (5), afterwards (4), almost before (1), an average of six months before (1), and (1), at least until (3), before (280), by then (1), even before (14), ever since (1), finally (10), five minutes before (1), fully eight months before (1), in the 3 1/2 years before (1), in the end (1), in turn (14), just before (5), just days before (1), just eight days before (1), later (88), later on (2), long before (2), next (6), only until (1), several months before (1), shortly afterward (1), shortly afterwards (1), shortly before (5), shortly thereafter (4), still (2), then (302), thereafter (5), till (3), two days before (1), two months before (1), two years before (1), ultimately (9), until (129), when (4), years before (1)	920
precedence/reason	by then (1), now that (1), then (1)	3
precedence/Restatement	ultimately (1)	1
precedence/result	and (1), finally (2), in turn (4), then (7), ultimately (2)	16
reason	apparently because (1), as (164), at least partly because (1), because (781), especially as (1), especially because (1), especially since (1), for (3), in large part because (1), in part because (11), insofar as (1), just because (6), largely because (12), mainly because (6), merely because (1), not because (3), not only because (1), now that (9), only because (3), particularly as (1), particularly because (2), particularly since (5), partly because (15), perhaps because (2), presumably because (1), primarily because (2), simply because (4), since (88), so (1), then (1), when (6)	1135
reason/Restatement	ultimately (1)	1

Table 11: Senses and their associated Explicit Connectives *(Cont. on next page)*

Sense	Explicit Connectives	Total
reason/succession	after (47), as (23), now that (4), once (7), one day after (1), only after (1), only when (1), reportedly after (1), since (10), when (64)	159
reason/Synchrony	as (141), especially as (2), now that (4), only when (1), particularly as (1), particularly when (1), when (37)	187
reason/TEMPORAL	as (1), when (1)	2
relevance	especially if (1), if (19)	20
relevance/succession	when (1)	1
Restatement	in fact (2), in other words (4), indeed (7), or (1), overall (3)	17
result	accordingly (5), and (38), as a result (77), but (1), consequently (10), hence (4), if only (1), in the end (1), in turn (3), largely as a result (1), now that (2), so (262), so that (31), then (4), thereby (12), therefore (26), thus (112)	590
result/succession	once (1), when (1)	2
result/Synchrony	as (2), when (2)	4
specification	also (1), and (1), as though (2), but (1), for example (2), if (1), in fact (36), in particular (9), in the end (1), in turn (1), indeed (32), much as (1), overall (4), rather (7), specifically (9)	108
specification/succession	especially after (1)	1
specification/Synchrony	especially when (1), when (1)	2
succession	18 months after (1), 25 years after (1), 29 years and 11 months to the day after (1), a day after (4), a few hours after (1), a few months after (1), a few weeks after (1), a month after (2), a week after (1), a year after (1), about a week after (1), about three weeks after (1), after (440), almost immediately after (1), as (3), as soon as (11), before (2), by then (3), earlier (14), eight months after (1), even after (5), ever since (6), five years after (1), four days after (1), immediately after (1), in the first 25 minutes after (1), in the meantime (1), just 15 days after (1), just a day after (1), just a month after (1), just after (2), just five months after (1), just minutes after (1), just when (1), less than a month after (1), long after (1), minutes after (2), months after (1), more than a year after (1), nearly a year and a half after (1), nearly two months after (1), once (67), one day after (2), only after (8), only three years after (1), only two weeks after (1), particularly after (1), previously (46), right after (1), seven years after (1), shortly after (9), since (72), since before (1), six years after (1), some time after (1), sometimes after (1), soon after (7), thereafter (1), three months after (1), two days after (1), two weeks after (2), until (5), when (156), within a year after (1), within minutes after (1), years after (1)	912

Table 12: Senses and their associated Explicit Connectives *(Cont. on next page)*

Sense	`Explicit` **Connectives**	**Total**
Synchrony	almost simultaneously (1), as (367), as long as (4), as soon as (8), at least not when (1), at least when (1), back when (1), by then (2), especially when (1), even as (7), even when (3), even while (3), in the meantime (10), in the meanwhile (1), just as (13), just as soon as (1), just when (5), meanwhile (25), now that (2), once (2), only when (1), simultaneously (5), since (1), still (1), then (5), until (1), usually when (1), when (463), while (151)	1087
TEMPORAL	when (3)	3
unreal_past	even if (2), if (50), if only (1), if..then (1)	54
unreal_present	even if (3), if (118), if only (1), then (1)	123

Table 13: Senses and their associated `Explicit` Connectives

Appendix C

This Appendix lists modified forms and variants of `Explicit` connectives (Tables 14-17). There are 100 distinct types of `Explicit` connectives (first column). The Total of the counts in the second and third columns is 18459.

`Explicit` **Connective**	**Modified forms and variants**	**Total**
accordingly	accordingly (5)	5
additionally	additionally (7)	7
after	18 months after (1), 25 years after (1), 29 years and 11 months to the day after (1), a day after (4), a few hours after (1), a few months after (1), a few weeks after (1), a month after (2), a week after (1), a year after (1), about a week after (1), about three weeks after (1), after (487), almost immediately after (1), eight months after (1), especially after (1), even after (8), five years after (1), four days after (1), immediately after (1), in the first 25 minutes after (1), just 15 days after (1), just a day after (1), just a month after (1), just after (2), just five months after (1), just minutes after (1), less than a month after (1), long after (1), minutes after (2), months after (1), more than a year after (1), nearly a year and a half after (1), nearly two months after (1), one day after (3), only after (9), only three years after (1), only two weeks after (1), particularly after (1), reportedly after (1), right after (1), seven years after (1), shortly after (9), six years after (1), some time after (1), sometimes after (1), soon after (7), three months after (1), two days after (1), two weeks after (2), within a year after (1), within minutes after (1), years after (1)	577
afterward	afterward (5), afterwards (4), shortly afterward (1), shortly afterwards (1)	11
also	also (1746)	1746
alternatively	alternatively (6)	6
although	although (328)	328
and	and (3000)	3000
as	as (711), especially as (3), even as (13), just as (14), particularly as (2)	743
as a result	as a result (77), largely as a result (1)	78
as an alternative	as an alternative (2)	2
as if	as if (16)	16
as long as	as long as (23), only as long as (1)	24
as soon as	as soon as (19), just as soon as (1)	20
as though	as though (5)	5
as well	as well (6)	6

Table 14: Modified forms and variants of `Explicit` connectives *(Cont. on next page)*

Explicit Connective	Modified forms and variants	Total
because	apparently because (1), at least partly because (1), because (783), especially because (1), in large part because (1), in part because (11), just because (8), largely because (12), mainly because (6), merely because (1), not because (3), not only because (1), only because (3), particularly because (2), partly because (15), perhaps because (2), presumably because (1), primarily because (2), simply because (4)	858
before	a day or two before (1), a decade before (1), a full five minutes before (1), a week before (1), about six months before (1), almost before (1), an average of six months before (1), before (282), even before (14), five minutes before (1), fully eight months before (1), in the 3 1/2 years before (1), just before (5), just days before (1), just eight days before (1), long before (2), several months before (1), shortly before (5), since before (1), two days before (1), two months before (1), two years before (1), years before (1)	326
before and after	before and after (1)	1
besides	besides (19)	19
but	but (3308)	3308
by comparison	by comparison (11)	11
by contrast	by contrast (27)	27
by then	by then (7)	7
consequently	consequently (10)	10
conversely	conversely (2)	2
earlier	earlier (15)	15
either..or	either..or (4)	4
else	else (1)	1
except	except (10)	10
finally	finally (32)	32
for	for (3)	3
for example	for example (196)	196
for instance	for instance (98)	98
further	further (9)	9
furthermore	furthermore (11)	11
hence	hence (4)	4
however	however (485)	485
if	especially if (6), even if (83), if (1111), if only (6), only if (13), particularly if (3), typically, if (1)	1223
if and when	if and when (3)	3
if..then	if..then (38)	38
in addition	in addition (165)	165
in contrast	in contrast (12)	12
in fact	in fact (82)	82
in other words	in other words (17)	17
in particular	in particular (15)	15
in short	in short (4)	4

Table 15: Modified forms and variants of Explicit connectives *(Cont. on next page)*

Explicit Connective	Modified forms and variants	Total
in sum	in sum (2)	2
in the end	in the end (9)	9
in turn	in turn (30)	30
indeed	indeed (104)	104
insofar as	insofar as (1)	1
instead	instead (112)	112
later	later (89), later on (2)	91
lest	lest (2)	2
likewise	likewise (8)	8
meantime	in the meantime (14), meantime (1)	15
meanwhile	in the meanwhile (1), meanwhile (192)	193
moreover	moreover (101)	101
much as	as much as (2), much as (3), so much as (1)	6
neither..nor	neither..nor (3)	3
nevertheless	nevertheless (44)	44
next	next (7)	7
nonetheless	nonetheless (27)	27
nor	nor (31)	31
now that	now that (22)	22
on the contrary	on the contrary (4)	4
on the one hand..on the other hand	on the one hand..on the other hand (1)	1
on the other hand	on the other hand (37)	37
once	once (84)	84
or	or (98)	98
otherwise	otherwise (24)	24
overall	overall (12)	12
plus	plus (1)	1
previously	previously (49)	49
rather	rather (17)	17
regardless	regardless (2)	2
separately	separately (74)	74
similarly	similarly (18)	18
simultaneously	almost simultaneously (1), simultaneously (5)	6
since	especially since (1), ever since (7), particularly since (5), since (171)	184
so	so (263)	263
so that	so that (31)	31
specifically	specifically (10)	10
still	even still (1), still (189)	190
then	even then (2), then (338)	340
thereafter	shortly thereafter (4), thereafter (7)	11
thereby	thereby (12)	12
therefore	therefore (26)	26
though	even though (95), though (225)	320

Table 16: Modified forms and variants of Explicit connectives *(Cont. on next page)*

Explicit Connective	Modified forms and variants	Total
thus	thus (112)	112
till	till (3)	3
ultimately	ultimately (18)	18
unless	unless (95)	95
until	at least until (3), just until (1), only until (1), until (157)	162
when	at least not when (1), at least when (2), back when (1), especially when (4), even when (14), except when (1), just when (6), only when (9), particularly when (1), usually when (1), when (949)	989
when and if	when and if (1)	1
whereas	whereas (5)	5
while	even while (3), while (778)	781
yet	yet (101)	101

Table 17: Modified forms and variants of Explicit connectives

Appendix D

This Appendix gives the distribution of `Implicit` connectives and their senses in PDTB-2.0., split across Tables 18-23. With multiple connectives being each counted separately (see Section 1.3 and Section 1.5 for discussion of how multiple connectives are annotated and represented) , the total number of tokens is 16224 (the Total of the counts in the second as well as the third column) (Note that when multiple connectives are not each counted separately, the total number of tokens is 16053, the number of tokens in which multiple connectives were provided being 171.) Multiple senses annotated for a single connective are shown as a distinct type of sense, the different senses separated by "/". See Figure 1 for the hierarchical classification of the senses shown here.

Implicit connective	**Senses**	**Total**
accordingly	Conjunction (1), result (82), result/specification (1), specification (1)	85
additionally	Conjunction (13)	13
after	Conjunction (1), reason/succession (9), succession (3)	13
afterwards	precedence (4)	4
also	Conjunction (456), Conjunction/result (1), conjunctive (1), List (2)	460
although	COMPARISON (18), Concession (4), Conjunction (2), contra-expectation (24), Contrast (100), Contrast/Pragmatic_Concession (1), Contrast/specification (2), EXPANSION/expectation (1), expectation (19), juxtaposition (27), List (1), opposition (2), specification (1)	202
and	Conjunction (891), Conjunction/juxtaposition (2), Conjunction/precedence (1), Conjunction/reason (1), Conjunction/Synchrony (1), Contrast (1), EXPANSION (3), Instantiation (1), juxtaposition (6), juxtaposition/List (7), List (346), opposition (1), precedence (1), result (2), specification (6), succession (1), Synchrony (1)	1272
as	Cause (1), Conjunction (11), Conjunction/Synchrony (3), EXPANSION (1), Instantiation/Synchrony (1), justification (11), justification/specification (3), reason (334), reason/specification (13), reason/Synchrony (3), result (1), specification (9), specification/Synchrony (3), Synchrony (17)	411
as a consequence	result (2)	2
as a matter of fact	Conjunction (1)	1
as a result	Conjunction/result (1), result (281), specification (1)	283
as it turns out	Conjunction (1), result (1)	2
at that time	Synchrony (1)	1
at the same time	COMPARISON (1), Conjunction (1), Conjunction/Synchrony (2), Contrast/Synchrony (2), Synchrony (4)	10
at the time	Conjunction/Synchrony (1), Synchrony (21)	22

Table 18: `Implicit` connectives and their senses *(Cont. on next page)*

Implicit connective	Senses	Total
because	Conjunction/reason (2), Contrast/reason (1), justification (31), justification/reason (6), List/reason (1), reason (1830), reason/specification (14), specification (2)	1887
before	succession (5)	5
besides	COMPARISON/Conjunction (1), Conjunction (14), Conjunction/justification (1)	16
but	COMPARISON (56), COMPARISON/Conjunction (1), COMPARISON/reason (1), Conjunction (42), Conjunction/Pragmatic_Contrast (1), conjunctive (1), contra-expectation (66), Contrast (409), Contrast/precedence (1), EXPANSION (3), expectation (2), hypothetical (1), juxtaposition (113), opposition (39), opposition/precedence (1), Pragmatic_Contrast (1), result (4), specification (1)	743
by comparison	COMPARISON (1), Conjunction (2), Conjunction/Contrast (1), Conjunction/juxtaposition (1), contra-expectation (1), Contrast (46), EXPANSION (1), juxtaposition (137), List/opposition (1), opposition (7)	198
by contrast	Alternative/juxtaposition (1), contra-expectation (1), Contrast (58), juxtaposition (65), juxtaposition/precedence (1), opposition (20)	146
consequently	Conjunction/result (1), result (190)	191
earlier	succession (28)	28
even though	COMPARISON (5), Concession (2), contra-expectation (7), contra-expectation/specification (1), Contrast (11), expectation (8), juxtaposition (2)	36
eventually	precedence (3), precedence/result (1), Restatement (1), specification (1)	6
ever since	precedence/result (1)	1
finally	Conjunction (1), juxtaposition/List (1), List (1), precedence (9), Restatement (1), result (1)	14
first	Conjunction (4), Instantiation (1), List (8), specification (13), succession (1)	27
for	reason (1)	1
for example	Conjunction (1), EXPANSION (2), Instantiation (733), Instantiation/justification (3), Instantiation/reason (1), reason (1), specification (102)	843
for instance	Contrast (1), Instantiation (581), Instantiation/justification (1), Instantiation/reason (1), List (1), specification (43)	628
for one	Instantiation (1)	1
for one thing	Conjunction (1), Instantiation (9), Instantiation/justification (1), Instantiation/reason (4), reason (1), reason/specification (2), specification (8)	26
further	Conjunction (57), specification (1)	58
furthermore	Conjunction (341), CONTINGENCY (1), List (1), result (1), specification (2)	346
hence	result (13)	13

Table 19: `Implicit` connectives and their senses *(Cont. on next page)*

Implicit connective	Senses	Total
however	COMPARISON (67), Conjunction (13), Conjunction/Contrast (1), Conjunction/Pragmatic_Contrast (1), contra-expectation (61), contra-expectation/specification (1), Contrast (413), Contrast/specification (1), EXPANSION (1), expectation (1), juxtaposition (120), opposition (27), Pragmatic_Contrast/specification (1), specification (1)	709
in addition	Conjunction (251), List (2), specification (1)	254
in comparison	juxtaposition (2)	2
in contrast	Contrast (8), juxtaposition (7), opposition (1)	16
in fact	Conjunction (436), Conjunction/CONTINGENCY (1), Conjunction/justification (1), Conjunction/juxtaposition (2), Conjunction/precedence (1), Conjunction/reason (2), Conjunction/result (2), contra-expectation (1), Contrast (2), Contrast/Instantiation (1), equivalence (5), EXPANSION (27), generalization (6), Instantiation (16), justification (7), justification/specification (1), juxtaposition (4), juxtaposition/specification (1), List (1), reason (2), reason/specification (1), Restatement (5), result (3), result/specification (1), specification (330)	859
in other words	Conjunction (3), CONTINGENCY/EXPANSION (1), equivalence (169), equivalence/reason (1), EXPANSION (3), generalization (25), reason (1), Restatement (33), result (1), specification (15)	252
in particular	Conjunction (1), Conjunction/juxtaposition (1), EXPANSION (2), Instantiation (54), justification/specification (1), Restatement (1), specification (570)	630
in response	Conjunction (1)	1
in return	Conjunction (1)	1
in short	Conjunction (6), equivalence (22), EXPANSION (7), generalization (75), reason (2), Restatement (22), result (1), specification (19)	154
in sum	Conjunction (6), equivalence (4), EXPANSION (1), generalization (22), Restatement (6), specification (1)	40
in summary	generalization (1)	1
in the end	Conjunction (19), equivalence (1), EXPANSION (8), generalization (3), precedence (2), precedence/result (3), Restatement (10), result (3), specification (4)	53
in the meantime	Synchrony (1)	1
in turn	Conjunction (9), Conjunction/contra-expectation (1), Conjunction/opposition (1), Conjunction/precedence (3), Conjunction/result (3), EXPANSION (1), precedence (5), result (3)	26
inasmuch as	justification (4), reason (12), reason/specification (1)	17
incidentally	Conjunction (1)	1

Table 20: `Implicit` connectives and their senses *(Cont. on next page)*

Implicit connective	Senses	Total
indeed	Conjunction (117), Conjunction/CONTINGENCY (1), Conjunction/justification (2), equivalence (37), equivalence/reason (1), EXPANSION (14), EXPANSION/reason (1), generalization (22), Instantiation (10), justification (3), reason (3), reason/specification (2), Restatement (66), result (1), specification (112)	392
insofar as	reason (2), reason/specification (1)	3
instead	Alternative (1), Alternative/chosen_alternative (1), chosen_alternative (107), Conjunction (1), Contrast (28), Exception (1), juxtaposition (1), opposition (1), specification (1)	142
later	Conjunction/precedence (2), precedence (11)	13
likewise	Conjunction (18)	18
meanwhile	Conjunction (81), Conjunction/Contrast (2), Conjunction/juxtaposition (6), Conjunction/Synchrony (75), Conjunction/TEMPORAL (1), Contrast (1), Contrast/Synchrony (3), EXPANSION (3), juxtaposition (7), juxtaposition/List (5), juxtaposition/Synchrony (6), List (2), List/Synchrony (3), precedence (1), Synchrony (34)	230
moreover	Conjunction (89), specification (1)	90
nevertheless	Conjunction (1), contra-expectation (14), Contrast (11), juxtaposition (2)	28
next	precedence (8)	8
nonetheless	contra-expectation (3), Contrast (1), juxtaposition (3)	7
now	precedence (1)	1
on the contrary	COMPARISON (1), Contrast (6), juxtaposition (1), opposition (2), specification (1)	11
on the one hand	Conjunction (1), Instantiation (2), specification (2)	5
on the other hand	COMPARISON (1), Contrast (16), juxtaposition (15), opposition (5)	37
on the whole	Conjunction (10), EXPANSION (2), generalization (18), Restatement (12), specification (8)	50
or	Alternative (1), conjunctive (8), equivalence (1), specification (1)	11
overall	Conjunction (11), Conjunction/result (1), EXPANSION (1), generalization (6), Restatement (7), specification (4)	30
particularly	specification (4)	4
plus	Conjunction (5), Conjunction/reason (1)	6
previously	precedence (2), succession (110)	112
rather	Alternative (1), chosen_alternative (63), Contrast (29), Contrast/EXPANSION (1), equivalence (1), Exception (1), generalization (1), juxtaposition (5), opposition (3), specification (30)	135
regardless	Contrast (2)	2
second	List (2)	2
separately	Conjunction (3)	3

Table 21: `Implicit` connectives and their senses *(Cont. on next page)*

Implicit connective	Senses	Total
similarly	Conjunction (65), EXPANSION (1)	66
simultaneously	Conjunction/Synchrony (1), Synchrony (2)	3
since	EXPANSION/justification (1), justification (4), reason (206), specification (2)	213
since then	precedence (5)	5
so	Conjunction (3), Conjunction/result (2), generalization (3), justification (1), precedence/result (2), reason (1), relevance (1), result (787), specification (1)	801
so far	EXPANSION (1)	1
so that	result (2)	2
soon	precedence (1)	1
specifically	Conjunction (6), equivalence (4), EXPANSION (1), generalization (1), Instantiation (24), Restatement (1), specification (1129)	1166
still	COMPARISON (1), Conjunction/Contrast (1), contra-expectation (3), Contrast (3), juxtaposition (2), opposition (2)	12
subsequently	EXPANSION (2), precedence (43), precedence/result (2), succession (1)	48
that is	Conjunction (2), equivalence (28), EXPANSION (2), generalization (6), reason (1), Restatement (43), result (3), specification (34)	119
then	Conjunction (3), Conjunction/precedence (2), List/precedence (2), opposition/precedence (1), precedence (396), precedence/result (4), precedence/specification (1), result (1), result/specification (1), specification (3), Synchrony (1)	415
thereafter	precedence (2)	2
therefore	result (113), result/specification (1)	114
third	List (1)	1
though	COMPARISON (1), Contrast (3), juxtaposition (1)	5
thus	justification (1), result (176), specification (2)	179
to this end	result (1)	1
ultimately	Conjunction (1), Conjunction/precedence (2), EXPANSION (1), EXPANSION/precedence (1), generalization (2), precedence (3), Restatement (6), result (3), specification (3)	22
what's more	Conjunction (1)	1
when	precedence (2), specification (1), specification/Synchrony (2), succession (3), Synchrony (10)	18
whereas	Conjunction (1), Conjunction/Contrast (1), Contrast (19), juxtaposition (40), juxtaposition/List (1), juxtaposition/precedence (1), opposition (11)	74

Table 22: `Implicit` connectives and their senses *(Cont. on next page)*

Implicit connective	Senses	Total
while	COMPARISON (1), Conjunction (336), Conjunction/Contrast (2), Conjunction/juxtaposition (10), Conjunction/opposition (2), Conjunction/precedence (3), Conjunction/Synchrony (31), contra-expectation (1), Contrast (44), Contrast/TEMPORAL (1), EXPANSION (1), juxtaposition (118), juxtaposition/List (3), juxtaposition/precedence (1), juxtaposition/Synchrony (12), List (6), List/Synchrony (2), opposition (15), opposition/Synchrony (1), Synchrony (7)	597
yet	COMPARISON (1), Conjunction (1), contra-expectation (1), Contrast (2), juxtaposition (1)	6

Table 23: `Implicit` connectives and their senses

Appendix E

This Appendix gives the distribution of all the distinct senses annotated for `Implicit` connectives (first column) along with their counts (third column), and the set of `Implicit` connectives for which the sense was annotated (second column) along with the counts of these connectives (in (parentheses). The full distribution is split across Tables 24-27. With multiple connectives each counted separately (see Section 1.3 and Section 1.5 for discussion of how multiple connectives are annotated and represented) , there are 84 distinct senses annotated for `Implicit` connectives, and the total number of `Implicit` connectives is 16224 (the Total of the counts in the second as well as the third column). (Note that when multiple connectives are not each counted separately, the total number of tokens is 16053, the number of tokens in which multiple connectives were provided being 171.) Multiple senses for a connective are shown as a distinct type of sense, the different senses separated by "/". See Figure 1 for the hierarchical classification of the senses shown here.

Sense	`Implicit` **connective**	**Total**
Alternative	instead (1), or (1), rather (1)	3
Alternative/chosen_alternative	instead (1)	1
Alternative/juxtaposition	by contrast (1)	1
Cause	as (1)	1
chosen_alternative	instead (107), rather (63)	170
COMPARISON	although (18), at the same time (1), but (56), by comparison (1), even though (5), however (67), on the contrary (1), on the other hand (1), still (1), though (1), while (1), yet (1)	154
COMPARISON/Conjunction	besides (1), but (1)	2
COMPARISON/reason	but (1)	1
Concession	although (4), even though (2)	6
Conjunction	accordingly (1), additionally (13), after (1), also (456), although (2), and (891), as (11), as a matter of fact (1), as it turns out (1), at the same time (1), besides (14), but (42), by comparison (2), finally (1), first (4), for example (1), for one thing (1), further (57), furthermore (341), however (13), in addition (251), in fact (436), in other words (3), in particular (1), in response (1), in return (1), in short (6), in sum (6), in the end (19), in turn (9), incidentally (1), indeed (117), instead (1), likewise (18), meanwhile (81), moreover (89), nevertheless (1), on the one hand (1), on the whole (10), overall (11), plus (5), separately (3), similarly (65), so (3), specifically (6), that is (2), then (3), ultimately (1), what's more (1), whereas (1), while (336), yet (1)	3344
Conjunction/CONTINGENCY	in fact (1), indeed (1)	2
Conjunction/contra-expectation	in turn (1)	1
Conjunction/Contrast	by comparison (1), however (1), meanwhile (2), still (1), whereas (1), while (2)	8

Table 24: Senses and their their associated `Implicit` connectives *(Cont. on next page)*

Sense	`Implicit` connective	Total
Conjunction/justification	besides (1), in fact (1), indeed (2)	4
Conjunction/juxtaposition	and (2), by comparison (1), in fact (2), in particular (1), meanwhile (6), while (10)	22
Conjunction/opposition	in turn (1), while (2)	3
Conjunction/Pragmatic_Contrast	but (1), however (1)	2
Conjunction/precedence	and (1), in fact (1), in turn (3), later (2), then (2), ultimately (2), while (3)	14
Conjunction/reason	and (1), because (2), in fact (2), plus (1)	6
Conjunction/result	also (1), as a result (1), consequently (1), in fact (2), in turn (3), overall (1), so (2)	11
Conjunction/Synchrony	and (1), as (3), at the same time (2), at the time (1), meanwhile (75), simultaneously (1), while (31)	114
Conjunction/TEMPORAL	meanwhile (1)	1
conjunctive	also (1), but (1), or (8)	10
CONTINGENCY	furthermore (1)	1
CONTINGENCY/EXPANSION	in other words (1)	1
contra-expectation	although (24), but (66), by comparison (1), by contrast (1), even though (7), however (61), in fact (1), nevertheless (14), nonetheless (3), still (3), while (1), yet (1)	183
contra-expectation/specification	even though (1), however (1)	2
Contrast	although (100), and (1), but (409), by comparison (46), by contrast (58), even though (11), for instance (1), however (413), in contrast (8), in fact (2), instead (28), meanwhile (1), nevertheless (11), nonetheless (1), on the contrary (6), on the other hand (16), rather (29), regardless (2), still (3), though (3), whereas (19), while (44), yet (2)	1214
Contrast/EXPANSION	rather (1)	1
Contrast/Instantiation	in fact (1)	1
Contrast/Pragmatic_Concession	although (1)	1
Contrast/precedence	but (1)	1
Contrast/reason	because (1)	1
Contrast/specification	although (2), however (1)	3
Contrast/Synchrony	at the same time (2), meanwhile (3)	5
Contrast/TEMPORAL	while (1)	1
equivalence	in fact (5), in other words (169), in short (22), in sum (4), in the end (1), indeed (37), or (1), rather (1), specifically (4), that is (28)	272
equivalence/reason	in other words (1), indeed (1)	2
Exception	instead (1), rather (1)	2
EXPANSION	and (3), as (1), but (3), by comparison (1), for example (2), however (1), in fact (27), in other words (3), in particular (2), in short (7), in sum (1), in the end (8), in turn (1), indeed (14), meanwhile (3), on the whole (2), overall (1), similarly (1), so far (1), specifically (1), subsequently (2), that is (2), ultimately (1), while (1)	89

Table 25: Senses and their their associated `Implicit` connectives *(Cont. on next page)*

Sense	Implicit connective	Total
EXPANSION/expectation	although (1)	1
EXPANSION/justification	since (1)	1
EXPANSION/precedence	ultimately (1)	1
EXPANSION/reason	indeed (1)	1
expectation	although (19), but (2), even though (8), however (1)	30
generalization	in fact (6), in other words (25), in short (75), in sum (22), in summary (1), in the end (3), indeed (22), on the whole (18), overall (6), rather (1), so (3), specifically (1), that is (6), ultimately (2)	191
hypothetical	but (1)	1
Instantiation	and (1), first (1), for example (733), for instance (581), for one (1), for one thing (9), in fact (16), in particular (54), indeed (10), on the one hand (2), specifically (24)	1432
Instantiation/justification	for example (3), for instance (1), for one thing (1)	5
Instantiation/reason	for example (1), for instance (1), for one thing (4)	6
Instantiation/Synchrony	as (1)	1
justification	as (11), because (31), in fact (7), inasmuch as (4), indeed (3), since (4), so (1), thus (1)	62
justification/reason	because (6)	6
justification/specification	as (3), in fact (1), in particular (1)	5
juxtaposition	although (27), and (6), but (113), by comparison (137), by contrast (65), even though (2), however (120), in comparison (2), in contrast (7), in fact (4), instead (1), meanwhile (7), nevertheless (2), nonetheless (3), on the contrary (1), on the other hand (15), rather (5), still (2), though (1), whereas (40), while (118), yet (1)	679
juxtaposition/List	and (7), finally (1), meanwhile (5), whereas (1), while (3)	17
juxtaposition/precedence	by contrast (1), whereas (1), while (1)	3
juxtaposition/specification	in fact (1)	1
juxtaposition/Synchrony	meanwhile (6), while (12)	18
List	also (2), although (1), and (346), finally (1), first (8), for instance (1), furthermore (1), in addition (2), in fact (1), meanwhile (2), second (2), third (1), while (6)	374
List/opposition	by comparison (1)	1
List/precedence	then (2)	2
List/reason	because (1)	1
List/Synchrony	meanwhile (3), while (2)	5
opposition	although (2), and (1), but (39), by comparison (7), by contrast (20), however (27), in contrast (1), instead (1), on the contrary (2), on the other hand (5), rather (3), still (2), whereas (11), while (15)	136
opposition/precedence	but (1), then (1)	2
opposition/Synchrony	while (1)	1
Pragmatic_Contrast	but (1)	1
Pragmatic_Contrast/specification	however (1)	1

Table 26: Senses and their their associated Implicit connectives *(Cont. on next page)*

Sense	`Implicit` connective	Total
precedence	afterwards (4), and (1), eventually (3), finally (9), in the end (2), in turn (5), later (11), meanwhile (1), next (8), now (1), previously (2), since then (5), soon (1), subsequently (43), then (396), thereafter (2), ultimately (3), when (2)	499
precedence/result	eventually (1), ever since (1), in the end (3), so (2), subsequently (2), then (4)	13
precedence/specification	then (1)	1
reason	as (334), because (1830), for (1), for example (1), for one thing (1), in fact (2), in other words (1), in short (2), inasmuch as (12), indeed (3), insofar as (2), since (206), so (1), that is (1)	2397
reason/specification	as (13), because (14), for one thing (2), in fact (1), inasmuch as (1), indeed (2), insofar as (1)	34
reason/succession	after (9)	9
reason/Synchrony	as (3)	3
relevance	so (1)	1
Restatement	eventually (1), finally (1), in fact (5), in other words (33), in particular (1), in short (22), in sum (6), in the end (10), indeed (66), on the whole (12), overall (7), specifically (1), that is (43), ultimately (6)	214
result	accordingly (82), and (2), as (1), as a consequence (2), as a result (281), as it turns out (1), but (4), consequently (190), finally (1), furthermore (1), hence (13), in fact (3), in other words (1), in short (1), in the end (3), in turn (3), indeed (1), so (787), so that (2), that is (3), then (1), therefore (113), thus (176), to this end (1), ultimately (3)	1676
result/specification	accordingly (1), in fact (1), then (1), therefore (1)	4
specification	accordingly (1), although (1), and (6), as (9), as a result (1), because (2), but (1), eventually (1), first (13), for example (102), for instance (43), for one thing (8), further (1), furthermore (2), however (1), in addition (1), in fact (330), in other words (15), in particular (570), in short (19), in sum (1), in the end (4), indeed (112), instead (1), moreover (1), on the contrary (1), on the one hand (2), on the whole (8), or (1), overall (4), particularly (4), rather (30), since (2), so (1), specifically (1129), that is (34), then (3), thus (2), ultimately (3), when (1)	2471
specification/Synchrony	as (3), when (2)	5
succession	after (3), and (1), before (5), earlier (28), first (1), previously (110), subsequently (1), when (3)	152
Synchrony	and (1), as (17), at that time (1), at the same time (4), at the time (21), in the meantime (1), meanwhile (34), simultaneously (2), then (1), when (10), while (7)	99

Table 27: Senses and their their associated `Implicit` connectives

Appendix F

This Apppendix gives the distribution of the 28 distinct types of `Altlex` senses in PDTB-2.0., along with their counts (Table 28). There are a total of 624 Altlex instances. Multiple senses annotated for an `AltLex` relation are shown as a distinct type of sense, the different senses separated by "/". See Figure 1 for the hierarchical classification of the senses shown here.

Altlex Sense	**Total**
contra-expectation	4
expectation	1
Contrast	11
juxtaposition	27
juxtaposition/succession	1
opposition	1
Contrast/disjunctive	1
Conjunction/Pragmatic_Contrast	1
reason	99
Instantiation/reason	1
reason/specification	3
result	170
precedence/result	1
general	2
justification	1
EXPANSION	3
Conjunction	111
Exception	1
Instantiation	37
List	1
Restatement	6
equivalence	5
generalization	12
specification	38
specification/succession	1
precedence	48
succession	18
Synchrony	19

Table 28: `Altlex` senses

Appendix G

This Appendix gives the distribution of the distinct attribution feature value sets annotated for `Explicit` connectives and their arguments. The full distribution is split across Tables 29-31. There are a total of 116 distinct feature value sets recorded (the Total of the fourth column).

Rel	**Arg1**	**Arg2**	**Total**
Arb.Comm.Null.Indet	Inh.Null.Null.Null	Inh.Null.Null.Null	1
Arb.Comm.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	11
Arb.Ctrl.Null.Indet	Inh.Null.Null.Null	Inh.Null.Null.Null	4
Arb.Ctrl.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	11
Arb.Ftv.Null.Indet	Inh.Null.Null.Null	Inh.Null.Null.Null	4
Arb.Ftv.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	7
Arb.PAtt.Null.Indet	Inh.Null.Null.Null	Inh.Null.Null.Null	4
Arb.PAtt.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	21
Ot.Comm.Null.Indet	Inh.Null.Null.Null	Inh.Null.Null.Null	31
Ot.Comm.Null.Null	Arb.Ctrl.Null.Null	Inh.Null.Null.Null	1
Ot.Comm.Null.Null	Arb.PAtt.Null.Null	Inh.Null.Null.Null	1
Ot.Comm.Null.Null	Inh.Null.Null.Indet	Inh.Null.Null.Null	1
Ot.Comm.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	3617
Ot.Comm.Null.Null	Inh.Null.Null.Null	Ot.Comm.Null.Null	3
Ot.Comm.Null.Null	Inh.Null.Null.Null	Ot.Ftv.Null.Null	1
Ot.Comm.Null.Null	Inh.Null.Null.Null	Ot.PAtt.Neg.Null	2
Ot.Comm.Null.Null	Inh.Null.Null.Null	Ot.PAtt.Null.Null	5
Ot.Comm.Null.Null	Ot.Comm.Neg.Null	Inh.Null.Null.Null	1
Ot.Comm.Null.Null	Ot.Comm.Null.Null	Inh.Null.Null.Null	283
Ot.Comm.Null.Null	Ot.Comm.Null.Null	Ot.PAtt.Neg.Null	1
Ot.Comm.Null.Null	Ot.Comm.Null.Null	Ot.PAtt.Null.Null	1
Ot.Comm.Null.Null	Ot.Ctrl.Neg.Null	Inh.Null.Null.Null	1
Ot.Comm.Null.Null	Ot.Ctrl.Null.Indet	Inh.Null.Null.Null	2
Ot.Comm.Null.Null	Ot.Ctrl.Null.Null	Inh.Null.Null.Null	9
Ot.Comm.Null.Null	Ot.Ctrl.Null.Null	Ot.Ctrl.Null.Null	1
Ot.Comm.Null.Null	Ot.Ftv.Null.Null	Inh.Null.Null.Null	8
Ot.Comm.Null.Null	Ot.PAtt.Neg.Null	Inh.Null.Null.Null	6
Ot.Comm.Null.Null	Ot.PAtt.Null.Null	Inh.Null.Null.Null	28
Ot.Comm.Null.Null	Ot.PAtt.Null.Null	Ot.PAtt.Null.Null	1
Ot.Comm.Null.Null	Wr.Comm.Null.Null	Inh.Null.Null.Null	104
Ot.Ctrl.Null.Indet	Inh.Null.Null.Null	Inh.Null.Null.Null	21
Ot.Ctrl.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	209
Ot.Ctrl.Null.Null	Ot.Ctrl.Null.Null	Inh.Null.Null.Null	1
Ot.Ctrl.Null.Null	Wr.Comm.Null.Null	Inh.Null.Null.Null	1
Ot.Ftv.Null.Indet	Inh.Null.Null.Null	Inh.Null.Null.Null	5
Ot.Ftv.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	58
Ot.Ftv.Null.Null	Inh.Null.Null.Null	Ot.PAtt.Null.Null	1
Ot.Ftv.Null.Null	Ot.PAtt.Null.Null	Inh.Null.Null.Null	1
Ot.Ftv.Null.Null	Wr.Comm.Null.Null	Inh.Null.Null.Null	2

Table 29: Attribution features of `Explicit` Connectives and their arguments *(Cont. on next page)*

Rel	Arg1	Arg2	Total
Ot.PAtt.Null.Indet	Inh.Null.Null.Null	Inh.Null.Null.Null	12
Ot.PAtt.Null.Null	Inh.Null.Neg.Null	Inh.Null.Null.Null	3
Ot.PAtt.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	186
Ot.PAtt.Null.Null	Ot.Comm.Null.Null	Inh.Null.Null.Null	6
Ot.PAtt.Null.Null	Ot.PAtt.Null.Null	Inh.Null.Null.Null	1
Ot.PAtt.Null.Null	Wr.Comm.Null.Null	Inh.Null.Null.Null	3
Wr.Comm.Null.Indet	Inh.Null.Null.Null	Inh.Null.Null.Null	2
Wr.Comm.Null.Null	Arb.Comm.Null.Null	Inh.Null.Null.Null	1
Wr.Comm.Null.Null	Arb.Ctrl.Null.Null	Inh.Null.Null.Null	1
Wr.Comm.Null.Null	Arb.Ctrl.Null.Null	Ot.Comm.Null.Null	1
Wr.Comm.Null.Null	Arb.Ftv.Null.Null	Inh.Null.Null.Null	3
Wr.Comm.Null.Null	Arb.PAtt.Neg.Null	Inh.Null.Null.Null	1
Wr.Comm.Null.Null	Arb.PAtt.Null.Indet	Inh.Null.Null.Null	1
Wr.Comm.Null.Null	Arb.PAtt.Null.Indet	Ot.Comm.Null.Null	1
Wr.Comm.Null.Null	Arb.PAtt.Null.Null	Inh.Null.Null.Null	4
Wr.Comm.Null.Null	Arb.PAtt.Null.Null	Ot.Comm.Null.Null	3
Wr.Comm.Null.Null	Inh.Null.Null.Null	Arb.Comm.Null.Null	2
Wr.Comm.Null.Null	Inh.Null.Null.Null	Arb.Ftv.Null.Null	1
Wr.Comm.Null.Null	Inh.Null.Null.Null	Arb.PAtt.Null.Null	2
Wr.Comm.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	12006
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.Comm.Neg.Null	1
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.Comm.Null.Null	402
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.Ctrl.Null.Null	5
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.Ftv.Null.Null	4
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.PAtt.Neg.Null	1
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.PAtt.Null.Indet	1
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.PAtt.Null.Null	20
Wr.Comm.Null.Null	Inh.Null.Null.Null	Wr.Ftv.Null.Null	1
Wr.Comm.Null.Null	Inh.Null.Null.Null	Wr.PAtt.Null.Indet	1
Wr.Comm.Null.Null	Inh.Null.Null.Null	Wr.PAtt.Null.Null	4
Wr.Comm.Null.Null	Ot.Comm.Null.Indet	Ot.Comm.Null.Null	3
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Arb.Comm.Null.Null	2
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Arb.PAtt.Null.Null	3
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Inh.Null.Null.Null	749
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.Comm.Neg.Null	1
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.Comm.Null.Null	329
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.Ctrl.Null.Null	8
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.Ftv.Null.Indet	1
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.Ftv.Null.Null	4
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.PAtt.Neg.Null	5
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.PAtt.Null.Null	23
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Wr.Comm.Null.Null	2
Wr.Comm.Null.Null	Ot.Ctrl.Neg.Null	Inh.Null.Null.Null	1
Wr.Comm.Null.Null	Ot.Ctrl.Null.Indet	Inh.Null.Null.Null	7

Table 30: Attribution features of `Explicit` Connectives and their arguments *(Cont. on next page)*

Rel	Arg1	Arg2	Total
Wr.Comm.Null.Null	Ot.Ctrl.Null.Indet	Ot.Comm.Null.Null	2
Wr.Comm.Null.Null	Ot.Ctrl.Null.Indet	Ot.Ctrl.Null.Null	1
Wr.Comm.Null.Null	Ot.Ctrl.Null.Null	Inh.Null.Null.Null	37
Wr.Comm.Null.Null	Ot.Ctrl.Null.Null	Ot.Comm.Null.Null	7
Wr.Comm.Null.Null	Ot.Ctrl.Null.Null	Ot.Ctrl.Null.Indet	1
Wr.Comm.Null.Null	Ot.Ctrl.Null.Null	Ot.Ctrl.Null.Null	2
Wr.Comm.Null.Null	Ot.Ctrl.Null.Null	Ot.PAtt.Null.Null	1
Wr.Comm.Null.Null	Ot.Ftv.Neg.Indet	Inh.Null.Null.Null	1
Wr.Comm.Null.Null	Ot.Ftv.Null.Null	Inh.Null.Null.Null	14
Wr.Comm.Null.Null	Ot.Ftv.Null.Null	Ot.Comm.Null.Null	6
Wr.Comm.Null.Null	Ot.Ftv.Null.Null	Ot.Ftv.Null.Null	1
Wr.Comm.Null.Null	Ot.Ftv.Null.Null	Ot.PAtt.Null.Null	1
Wr.Comm.Null.Null	Ot.PAtt.Neg.Null	Inh.Null.Null.Null	3
Wr.Comm.Null.Null	Ot.PAtt.Neg.Null	Ot.Comm.Null.Null	3
Wr.Comm.Null.Null	Ot.PAtt.Neg.Null	Ot.PAtt.Null.Null	1
Wr.Comm.Null.Null	Ot.PAtt.Null.Indet	Inh.Null.Null.Null	3
Wr.Comm.Null.Null	Ot.PAtt.Null.Null	Inh.Null.Null.Null	43
Wr.Comm.Null.Null	Ot.PAtt.Null.Null	Ot.Comm.Null.Null	20
Wr.Comm.Null.Null	Ot.PAtt.Null.Null	Ot.Ctrl.Null.Null	2
Wr.Comm.Null.Null	Ot.PAtt.Null.Null	Ot.PAtt.Neg.Null	1
Wr.Comm.Null.Null	Ot.PAtt.Null.Null	Ot.PAtt.Null.Null	3
Wr.Comm.Null.Null	Wr.Comm.Null.Null	Inh.Null.Null.Null	3
Wr.Comm.Null.Null	Wr.Ctrl.Null.Null	Inh.Null.Null.Null	1
Wr.Comm.Null.Null	Wr.Ftv.Null.Null	Inh.Null.Null.Null	1
Wr.Comm.Null.Null	Wr.PAtt.Neg.Null	Wr.PAtt.Null.Null	1
Wr.Comm.Null.Null	Wr.PAtt.Null.Null	Inh.Null.Null.Null	1
Wr.Ctrl.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	1
Wr.Ctrl.Null.Null	Ot.Ctrl.Null.Null	Inh.Null.Null.Null	1
Wr.Ftv.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	3
Wr.PAtt.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	4
Wr.PAtt.Null.Null	Wr.PAtt.Null.Null	Inh.Null.Null.Null	2

Table 31: Attribution features of `Explicit` Connectives and their arguments

Appendix H

This Appendix gives the distribution of the distinct attribution feature value sets annotated for `Implicit` connectives and `AltLex` relations, and their arguments. The full distribution is split across Tables 32-33. There are a total of 80 distinct feature value sets recorded (the Total of the fourth column).

Rel	**Arg1**	**Arg2**	**Total**
Arb.Comm.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	2
Ot.Comm.Null.Null	Arb.PAtt.Null.Null	Inh.Null.Null.Null	1
Ot.Comm.Null.Null	Inh.Null.Null.Null	Arb.PAtt.Null.Null	1
Ot.Comm.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	1469
Ot.Comm.Null.Null	Inh.Null.Null.Null	Ot.Comm.Null.Null	4
Ot.Comm.Null.Null	Inh.Null.Null.Null	Ot.Ctrl.Null.Null	1
Ot.Comm.Null.Null	Inh.Null.Null.Null	Ot.PAtt.Neg.Null	7
Ot.Comm.Null.Null	Inh.Null.Null.Null	Ot.PAtt.Null.Null	23
Ot.Comm.Null.Null	Ot.Comm.Neg.Null	Inh.Null.Null.Null	1
Ot.Comm.Null.Null	Ot.Comm.Null.Null	Inh.Null.Null.Indet	1
Ot.Comm.Null.Null	Ot.Comm.Null.Null	Inh.Null.Null.Null	306
Ot.Comm.Null.Null	Ot.Comm.Null.Null	Ot.Comm.Null.Null	3
Ot.Comm.Null.Null	Ot.Comm.Null.Null	Ot.PAtt.Neg.Null	1
Ot.Comm.Null.Null	Ot.Comm.Null.Null	Ot.PAtt.Null.Null	1
Ot.Comm.Null.Null	Ot.Ftv.Null.Indet	Inh.Null.Null.Null	1
Ot.Comm.Null.Null	Ot.Ftv.Null.Null	Inh.Null.Null.Null	9
Ot.Comm.Null.Null	Ot.PAtt.Neg.Null	Inh.Null.Null.Null	1
Ot.Comm.Null.Null	Ot.PAtt.Neg.Null	Ot.PAtt.Neg.Null	2
Ot.Comm.Null.Null	Ot.PAtt.Neg.Null	Ot.PAtt.Null.Null	1
Ot.Comm.Null.Null	Ot.PAtt.Null.Null	Inh.Null.Null.Null	25
Ot.Comm.Null.Null	Ot.PAtt.Null.Null	Ot.PAtt.Null.Null	6
Ot.Ctrl.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	3
Ot.Ftv.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	3
Ot.Ftv.Null.Null	Ot.Comm.Null.Null	Inh.Null.Null.Null	1
Ot.Ftv.Null.Null	Ot.Ftv.Null.Null	Inh.Null.Null.Null	1
Ot.PAtt.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	7
Ot.PAtt.Null.Null	Ot.Comm.Null.Null	Inh.Null.Null.Null	4
Ot.PAtt.Null.Null	Ot.PAtt.Null.Null	Inh.Null.Null.Null	1
Wr.Comm.Null.Null	Arb.Comm.Neg.Null	Inh.Null.Null.Null	1
Wr.Comm.Null.Null	Arb.Comm.Null.Null	Inh.Null.Null.Null	5
Wr.Comm.Null.Null	Arb.Comm.Null.Null	Ot.Comm.Null.Null	3
Wr.Comm.Null.Null	Arb.Ftv.Null.Null	Inh.Null.Null.Null	1
Wr.Comm.Null.Null	Arb.Ftv.Null.Null	Ot.Comm.Null.Null	1
Wr.Comm.Null.Null	Arb.PAtt.Neg.Null	Inh.Null.Null.Null	1
Wr.Comm.Null.Null	Arb.PAtt.Neg.Null	Ot.Comm.Null.Null	1
Wr.Comm.Null.Null	Arb.PAtt.Null.Indet	Inh.Null.Null.Null	1

Table 32: Attribution features of `Implicit` Connectives and `AltLex` relations, and their arguments
(Cont. on next page)

Rel	Arg1	Arg2	Total
Wr.Comm.Null.Null	Arb.PAtt.Null.Null	Arb.PAtt.Null.Null	1
Wr.Comm.Null.Null	Arb.PAtt.Null.Null	Inh.Null.Null.Null	8
Wr.Comm.Null.Null	Arb.PAtt.Null.Null	Ot.Comm.Null.Null	4
Wr.Comm.Null.Null	Inh.Null.Null.Null	Arb.Comm.Null.Null	6
Wr.Comm.Null.Null	Inh.Null.Null.Null	Arb.Ftv.Null.Null	1
Wr.Comm.Null.Null	Inh.Null.Null.Null	Arb.PAtt.Null.Indet	1
Wr.Comm.Null.Null	Inh.Null.Null.Null	Arb.PAtt.Null.Null	4
Wr.Comm.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	10342
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.Comm.Null.Null	1482
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.Ctrl.Null.Null	7
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.Ftv.Null.Null	21
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.PAtt.Neg.Null	5
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.PAtt.Null.Null	69
Wr.Comm.Null.Null	Inh.Null.Null.Null	Wr.Comm.Null.Null	1
Wr.Comm.Null.Null	Inh.Null.Null.Null	Wr.Ctrl.Null.Null	2
Wr.Comm.Null.Null	Ot.Comm.Neg.Null	Ot.Comm.Null.Null	6
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Arb.Comm.Null.Null	6
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Arb.Ftv.Null.Null	1
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Arb.PAtt.Null.Null	3
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Inh.Null.Null.Null	846
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.Comm.Null.Indet	1
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.Comm.Null.Null	1060
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.Ctrl.Neg.Null	1
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.Ctrl.Null.Null	2
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.Ftv.Null.Null	9
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.PAtt.Neg.Null	7
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.PAtt.Null.Null	54
Wr.Comm.Null.Null	Ot.Ctrl.Null.Null	Inh.Null.Null.Null	6
Wr.Comm.Null.Null	Ot.Ctrl.Null.Null	Ot.Comm.Null.Null	3
Wr.Comm.Null.Null	Ot.Ctrl.Null.Null	Ot.PAtt.Null.Null	1
Wr.Comm.Null.Null	Ot.Ftv.Null.Null	Inh.Null.Null.Null	14
Wr.Comm.Null.Null	Ot.Ftv.Null.Null	Ot.Comm.Null.Null	23
Wr.Comm.Null.Null	Ot.Ftv.Null.Null	Ot.PAtt.Null.Null	2
Wr.Comm.Null.Null	Ot.PAtt.Neg.Null	Ot.Comm.Null.Null	4
Wr.Comm.Null.Null	Ot.PAtt.Neg.Null	Ot.PAtt.Neg.Null	1
Wr.Comm.Null.Null	Ot.PAtt.Neg.Null	Ot.PAtt.Null.Null	4
Wr.Comm.Null.Null	Ot.PAtt.Null.Null	Arb.PAtt.Null.Null	1
Wr.Comm.Null.Null	Ot.PAtt.Null.Null	Inh.Null.Null.Null	60
Wr.Comm.Null.Null	Ot.PAtt.Null.Null	Ot.Comm.Null.Null	58
Wr.Comm.Null.Null	Ot.PAtt.Null.Null	Ot.Ftv.Null.Null	2
Wr.Comm.Null.Null	Ot.PAtt.Null.Null	Ot.PAtt.Null.Null	17
Wr.Comm.Null.Null	Wr.Ctrl.Null.Null	Inh.Null.Null.Null	2
Wr.Comm.Null.Null	Wr.Ftv.Null.Null	Inh.Null.Null.Null	4
Wr.Comm.Null.Null	Wr.PAtt.Null.Null	Inh.Null.Null.Null	1

Table 33: Attribution features of `Implicit` Connectives and `AltLex` relations, and their arguments

References

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

Dinesh, N., Lee, A., Miltsakaki, E., Prasad, R., Joshi, A., and Webber, B. (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*, Ann Arbor, Michigan.

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

Giordano, L. and Schwind, C. (2004). Conditional logic of actions and causation. *Artificial Intelligence*, **157**(1-2), 239–279.

Grosz, B. J., Joshi, A. K., and Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, **21**(2), 203–225.

Hintikka, J. (1971). Semantics for propositional attitudes. In L. Linsky, editor, *Reference and Modality*, pages 145–167. Oxford.

Hirschberg, J. and Litman, D. J. (1987). Now let's talk about NOW: Identifying cue phrases intonationally. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, pages 163–171.

Horn, L. (1978). Remarks on neg-raising. In P. Cole, editor, *Syntax and Semantics 9: Pragmatics*. Academic Press, New York.

Huddleston, R. and Pullum, G. (2002). *The Cambridge Grammar of the English Language*. Cambridge Univ. Press, Cambridge, UK.

Iatridou, S. (2000). The grammatical ingredients of counterfactuality. *Linguistic Inquiry*, **31**(2).

Karttunen, L. (1971). Some observations on factivity. *Papers in Linguistics*, **4**, 55–69.

Kingsbury, P. and Palmer, M. (2002). From Treebank to Propbank. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC-2002)*, Las Palmas, Canary Islands, Spain.

Kiparsky, C. and Kiparsky, P. (1971). Fact. In D. D. Steinberg and L. A. Jakobovits, editors, *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*, pages 345–369. Cambridge University Press, Cambridge.

Knott, A., Oberlander, J., O'Donnell, M., and Mellish, C. (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. Benjamins, Amsterdam.

Lasnik, H. (1975). On the semantics of negation. In *Contemporary Research in Philosophical Logic and Linguistic Semantics*, pages 279–313. Dordrecht: D. Reidel.

Levin, B. (1993). *English verb classes and alternations : a preliminary investigation*. University of Chicago Press.

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

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

Miltsakaki, E., Creswell, C., Forbes, K., Joshi, A., and Webber, B. (2003). Anaphoric arguments of discourse connectives: Semantic properties of antecedents versus non-antecedents. In *Proceedings of the Computational Treatment of Anaphora Workshop, EACL 2003, Budapest*.

Miltsakaki, E., Prasad, R., Joshi, A., and Webber, B. (2004a). Annotating discourse connectives and their arguments. In *Proceedings of the HLT/NAACL Workshop on Frontiers in Corpus Annotation*, pages 9–16, Boston, MA.

Miltsakaki, E., Prasad, R., Joshi, A., and Webber, B. (2004b). The Penn Discourse Treebank. In *Proceedings of the 4th International Conference on Language Rescourses and Evaluation (LREC 2004)*, Lisbon, Portugal.

Miltsakaki, E., Dinesh, N., Prasad, R., Joshi, A., and Webber, B. (2005). Experiments on sense annotation and sense disambiguation of discourse connectives. In *Proceedings of the Fourth Workshop on Treebanks and Linguistic Theories (TLT2005)*, Barcelona, Spain.

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

Prasad, R., Miltsakaki, E., Joshi, A., and Webber, B. (2004). Annotation and data mining of the Penn Discourse Treebank. In *Proceedings of the ACL Workshop on Discourse Annotation*, pages 88–95, Barcelona, Spain.

Prasad, R., Joshi, A., Dinesh, N., Lee, A., Miltsakaki, E., and Webber, B. (2005). The Penn Discourse TreeBank as a resource for natural language generation. In *Proceedings of the Corpus Linguistics Workshop on Using Corpora for NLG*, Birmingham, U.K.

Prasad, R., Dinesh, N., Lee, A., Joshi, A., and Webber, B. (2006). Annotating attribution in the Penn Discourse TreeBank. In *Proceedings of the COLING/ACL Workshop on Sentiment and Subjectivity in Text*, pages 31–38.

Prasad, R., Dinesh, N., Lee, A., Joshi, A., and Webber, B. (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).

Riloff, E. and Wiebe, J. (2003). Learning extraction patterns for subjective expressions. In *Proceedings of the SIGDAT Conference on Empirical Methods in Natural Language Processing (EMNLP03)*, pages 105–112, Sapporo, Japan.

Sag, I. A. and Pollard, C. (1991). An integrated theory of complement control. *Language*, **67**(1), 63–113.

Webber, B. and Di Eugenio, B. (1990). Free adjuncts in natural language instructions. In *Proceedings of COLING90*, pages 395–400.

Webber, B. and Joshi, A. (1998). Anchoring a lexicalized tree-adjoining grammar for discourse. In M. Stede, L. Wanner, and E. Hovy, editors, *Discourse Relations and Discourse Markers: Proceedings of the Conference*, pages 86–92. Association for Computational Linguistics, Somerset, New Jersey.

Webber, B., Knott, A., and Joshi, A. (1999). Multiple discourse connectives in a lexicalized grammar for discourse. In *Proceedings of the Third International Workshop on Computational Semantics, Tilberg, The Netherlands.*, pages 309–325.

Webber, B., Joshi, A., Stone, M., and Knott, A. (2003). Anaphora and discourse structure. *Computational Linguistics*, **29**(4), 545–587.

Webber, B., Joshi, A., Miltsakaki, E., Prasad, R., Dinesh, N., Lee, A., and Forbes, K. (2005). A short introduction to the PDTB. In *Copenhagen Working Papers in Language and Speech Processing*.

Wiebe, J., Wilson, T., Bruce, R., Bell, M., and Martin, M. (2004). Learning subjective language. *Computational Linguistics*, **30**(3), 277–308.

Wiebe, J., Wilson, T., , and Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, **1**(2).