

# The Penn Discourse Treebank 3.0 Annotation Manual

Bonnie Webber      Rashmi Prasad      Alan Lee      Aravind Joshi  
University of Edinburgh      Interactions, LLC      University of Pennsylvania

## In Memoriam

Our colleague and friend, and the guiding light of the Penn Discourse Treebank, Professor Aravind K. Joshi, died on 31 December 2017, before we were able to complete this new version of the corpus (the PDTB-3). We miss him more than words can say and dedicate the corpus to his memory.

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>What’s New in the PDTB-3?</b>	<b>4</b>
2.1	New Senses . . . . .	5
2.2	Multi-sense connectives whose sense depends on their PoS . . . . .	6
2.3	Discontinuous Connectives . . . . .	6
2.4	Paired Connectives . . . . .	7
2.5	Explicit Connectives in Intra-Sentential Discourse Relations . . . . .	7
2.6	Spans that can be analyzed as one connective or two . . . . .	7
2.7	New Relation Types . . . . .	9
2.7.1	Hypophora . . . . .	9
2.7.2	AltLexC . . . . .	9
<b>3</b>	<b>What’s different in the PDTB-3?</b>	<b>10</b>
3.1	Labelling Arguments and their spans . . . . .	11
3.2	Identifying AltLex . . . . .	12
3.3	Attribution . . . . .	14
3.4	Multiple Relations between Arguments . . . . .	14
3.5	Annotating Contrast vs Concession . . . . .	14
3.6	Contents of the Conn field . . . . .	15
3.7	Adverbials removed as Explicit Discourse Connectives . . . . .	15
<b>4</b>	<b>PDTB-3 Senses</b>	<b>16</b>
4.1	Differences between the PDTB-2 and PDTB-3 sense hierarchy . . . . .	16
4.1.1	Simplifying the relation hierarchy . . . . .	16
4.1.2	Augmenting the relation hierarchy . . . . .	18
4.2	Sense Classification . . . . .	18
4.2.1	Temporal Relations . . . . .	18
4.2.2	Contingency Relations . . . . .	19
4.2.3	Comparison Relations . . . . .	23
4.2.4	Expansion Relations . . . . .	25
4.3	Effect of Modifiers on Connectives: <i>Even</i> . . . . .	31
4.4	Effect of Explicit and Implicit Negation . . . . .	31
<b>5</b>	<b>Intra-sentential Discourse Annotation</b>	<b>32</b>
5.1	Free Adjuncts (FAs) . . . . .	33
5.1.1	Implicit connectives in FAs . . . . .	34

5.1.2	AltLex phrases in FAs . . . . .	35
5.2	Infinitival Clauses . . . . .	36
5.3	Other Subordinated Structures . . . . .	39
5.4	Conjoined Verb Phrases . . . . .	40
5.5	Implicit Conjoined Clauses . . . . .	43
5.5.1	Argument Span Annotation . . . . .	43
5.5.2	Sense Annotation . . . . .	44
5.6	Newly Observed Ambiguities in Intra-sentential Connectives . . . . .	45
5.6.1	Ambiguity of Concession Markers: Arg1-as-denier vs. Arg2-as-denier . . . . .	45
5.6.2	Ambiguity of <i>So</i> and <i>so that</i> : Purpose vs. Result . . . . .	46
<b>6</b>	<b>Inter-sentential Discourse Annotation</b>	<b>47</b>
6.1	Inter-sentential relations missing from the PDTB-2 . . . . .	47
6.2	Re-annotating PDTB-2 tokens due to changes in the sense-hierarchy . . . . .	47
<b>7</b>	<b>Consistency Checking</b>	<b>47</b>
<b>8</b>	<b>Implementation</b>	<b>48</b>
8.1	Format of Relation Tokens . . . . .	48
8.2	Provenance . . . . .	49
8.3	Annotation Process . . . . .	50
8.4	Overlapping Tokens . . . . .	50
8.5	Format of Multiple Relations . . . . .	52
<b>A</b>	<b>Explicit Connectives and their Senses</b>	<b>55</b>
<b>B</b>	<b>Explicit Connectives and their Modifiers</b>	<b>65</b>
<b>C</b>	<b>Implicit Relation Senses and Implicit Connectives</b>	<b>72</b>
<b>D</b>	<b>AltLex Senses</b>	<b>75</b>
<b>E</b>	<b>AltLexC Senses</b>	<b>76</b>
<b>F</b>	<b>Explicit Relations Linked together with Other Explicits</b>	<b>77</b>
<b>G</b>	<b>Explicit Relations Linked with Implicits</b>	<b>78</b>
<b>H</b>	<b>AltLex Relations Linked with Explicits</b>	<b>80</b>
<b>I</b>	<b>AltLex Relations Linked with Implicits</b>	<b>81</b>

# 1 Introduction

The last decade has seen growing interest in enabling language technology and psycholinguistics to move beyond the sentence, to what can be derived from larger units of text. This has led to greater interest in the properties of discourse. One such property is the coherence between clauses and sentences arising from low-level discourse relations. This level of meaning has been made overt through manual annotation in the Penn Discourse TreeBank (PDTB), developed with NSF support. Version 2.0. of the PDTB (Prasad et al., 2008), released in 2008, contains 40600 tokens of annotated relations, making it the largest such corpus available today. Largely because the PDTB was based on the simple idea that discourse relations are grounded in an identifiable set of explicit words or phrases (discourse connectives) or simply in the adjacency of two sentences, it has been taken up and used by many researchers in the NLP community and more recently, by researchers in psycholinguistics as well. It has also stimulated the development of similar resources in other languages (Chinese, Czech, Hindi, Modern Standard Arabic, Turkish and most recently, French) and domains (biomedical texts, conversational dialogues), the organization of community-level shared tasks on *shallow discourse parsing* (Xue et al., 2015, 2016), and a cross-lingual discourse annotation of parallel texts, the TED-MDB corpus (Zeyrek et al., 2018), to support both linguistic understanding of coherence in different languages and improvements in machine translation of discourse connectives.

Given only three years in which to develop guidelines, and annotate and release the PDTB, we knew that it would be incomplete. However, we felt it was important to deliver an annotated resource and learn from its use. This we have done: Use of the PDTB has made clear its value to the community, but also the extent of its incompleteness and the value for both language technology and psycholinguistics research of eliminating this gap.

This has led to our work on a new version of the corpus — the PDTB-3. This manual begins by summarizing what is new in the PDTB-3 and how it differs from what is in the PDTB-2. It is important to note that every token in the PDTB-3 is labelled with its *provenance* — indicating whether it is a copy of a token from the PDTB-2, a modified version of a PDTB-2 token, or new to the PDTB-3 (cf. Section 8.3). This provenance is designed to allow researchers to compare their earlier results on the PDTB-2 with results on an updated version of the corpus, as well as to produce new results based on the entire PDTB-3 corpus.

Finally, we acknowledge the following for their contributions to the PDTB-3: Sam Gibbon, Robin Malamud, and James Reed for all their work on corpus annotation, and Marta Andersson (Stockholm University) for the judgments of *so* done for her PhD thesis.

## 2 What’s New in the PDTB-3?

Simply put, the PDTB-3 is both bigger and better than the PDTB-2. With respect to size, the PDTB-3 contains ~13K more tokens annotated for discourse relations, for a total of 53631 tokens. With respect to quality, certain pairwise annotation decisions have been standardized (e.g., CONTRAST vs. CONCESSION) and applied to all previously, as well as newly, annotated tokens; certain difficult to annotate senses have been dropped in favor of ones that are easier to annotate; new senses have been added as more appropriate labels for already annotated tokens as well as for new tokens (Section 4); and a battery of *consistency checks* were done across the corpus, to ensure that similar tokens were either annotated in similar ways or shown to differ (Section 7). Similar tokens with different labels were corrected to bring

RelType	Intra-S		Diff	Inter-S		Diff
	PDTB-2	PDTB-3		PDTB-2	PDTB-3	
Explicit	11209	16905	5696	7243	7330	87
Implicit	531	6188	5657	15516	15544	28
AltLex	–	785	785	624	712	88
AltLexC	–	130	130	–	5	5
EntRel	53	358	305	5155	5172	17
Hypophora	–	4	4	–	142	142
NoRel	–	–	–	252	283	31
Total	11794	24369	–	28789	29818	–

Figure 1: Distribution of tokens annotated in the PDTB-2 and PDTB-3, within and across sentences. Explicit tokens have an explicit connective. AltLex and AltLexC tokens lack an explicit connective, but contain other phrasal or construction-based evidence for the relation that holds between the arguments. In EntRel tokens, a relation holds between an entity mentioned in Arg1 and the contents of Arg2. Hypophora tokens involve a question posed in Arg1 and an answer in Arg2. Relations that hold in Implicit tokens must be inferred, while NoRel indicates (of adjacent sentences within a paragraph) that no relation holds between them.

annotation in line across the corpus as a whole, and overlapping tokens were checked for consistency with each other. Note that this is not the same as assessing Inter-annotator Agreement (IAA), which only addresses annotator disagreement on individual tokens.

With respect to additional annotation, most are of intra-sentential (*Intra-S*) discourse relations. A token is considered *Intra-S* if it lies wholly within the projection of a top-level S-node in the Penn TreeBank. Otherwise, the token is considered inter-sentential (*Inter-S*). Newly annotated *Intra-S* tokens include relations between the conjuncts in conjoined verb phrases (Section 5.4) and conjoined clauses (Section 5.5), relations between free or headed adjuncts and the clauses they adjoin to (Section 5.1), relations between infinitival clauses and their matrix clause (Section 5.2), and relations between other subordinating structures and their matrix clauses (Section 5.3).

New annotation also includes explicitly marked question-response pairs called HYPOPHORA (Section 2.7.1). Most of these are *Inter-S*, except for four tokens within quoted speech that have been parsed as sentence-internal fragments. Also newly annotated are lexico-syntactic constructions that are unambiguous signals of particular discourse relations. These have been classed as ALTLEXC, (Section 2.7.2) to indicate that they are a type of AltLex construction, but also to allow them to be found on their own, separate from other AltLex tokens. A comparison of the distribution of Relation Types annotated in the PDTB-2 and PDTB-3 is given in Figure 1. The rest of this section provides brief descriptions of what’s new, described in greater detail later on.

## 2.1 New Senses

Annotating *Intra-S* discourse relations turned out to require senses that had not been required previously. In some cases, the senses were completely new; in some cases, the senses were negative versions of senses that had been noted in positive form, and in some cases, the senses involved reversing the arguments – e.g., where a sense used in the PDTB-2 involved **arg2** providing detail for *arg1*, what was now needed

was a sense in which *arg1* provides detail for *arg2*. These new senses are described in Section 4.

## 2.2 Multi-sense connectives whose sense depends on their PoS

Annotating *Intra-S* discourse relations also called attention to the fact that a discourse connective could have different senses, depending on its part-of-speech (PoS), a characteristic that had already been noted of discourse connectives in German (Scheffler and Stede, 2016).

Here are some English examples from the PDTB-3. While *however* was only seen as an adverbial in the PDTB-2, intra-sententially, it can also occur as a subordinator, as in Example 1.

- (1) *The 1987 crash was "a false alarm however you view it,"* says University of Chicago economist Victor Zarnowitz. [wsj\_2397]

(Those cases in which *however* appears as a subordinator simply with an adjective or adverb as in Example 2, have not been annotated in the PDTB-3 because they lack a verb.)

- (2) However risky the business, *it's brisk these days.* [wsj\_0569]

While as a discourse adverbial, the most common sense of *however* is ARG2-AS-DENIER, as a subordinator, its most common sense is ARG1-AS-DENIER — the reverse of ARG2-AS-DENIER.

While *since* was recognized as semantically ambiguous in the PDTB-2, *Intra-S* annotation showed it to have a PoS ambiguity as well — as a subordinating conjunction and as a discourse adverbial. As a subordinating conjunction, *since* can convey either REASON or TEMPORAL.SUCCESION (or both). As a discourse adverbial, however, it conveys the sense of PRECEDENCE (as in "He left for the Altacama Desert and hasn't been heard of since"), the reverse of SUCCESION.

Like *since*, *before* appears as both a subordinating conjunction and a discourse adverbial. As a subordinating conjunction, *before* conveys PRECEDENCE, while as an adverbial, it conveys SUCCESION — i.e., the reverse of PRECEDENCE.

Because Part-of-speech tags are not provided in the PDTB-3 corpus, a separate process of PoS-tagging can benefit discourse parsing in two ways: It can identify whether a word or phrase is being used as a discourse connective at all, and the PoS-tag it assigns can provide evidence for sense labelling.

## 2.3 Discontinuous Connectives

Although temporal subordinating conjunctions like *before*, *after*, *when* and *while* were annotated in the PDTB-2 both in "bare" form and with a temporal modifier such as "three days after", "a week before" (cf. Appendix B), they weren't annotated when combined with a temporal specifier such as "last year" or "a year", as in

- (3) And where were all our handsomely paid Indian Ocean allies last year when our convoys were being attacked? [wsj\_2112]
- (4) AMR's earnings decline comes a year after the parent company of American Airlines and the rest of the airline industry set profit records. [wsj\_1812]

While such tokens have now been annotated in the PDTB-3, they can introduce discontinuous connectives when the temporal subordinating conjunction has a modifier such as "particularly", as in

- (5) *These have been among the country's leading imports, particularly last year when there were shortages that led many traders to buy heavily and pay dearly.* [wsj\_1469]

In such cases, the modifier is taken to go with the conjunction, not with the temporal specifier.

## 2.4 Paired Connectives

In addition to the *paired connectives* that appear in the PDTB-2 (*either . . . or, neither . . . nor, both . . . and, not only . . . but also* and *on the one hand . . . on the other (hand)*), two additional paired connectives have been annotated in the PDTB-3 – *not just . . . but*, and *not only . . . but*.

## 2.5 Explicit Connectives in Intra-Sentential Discourse Relations

In the PDTB-2, explicit connectives included only coordinating and subordinating conjunctions and discourse adverbials. To more fully annotate *Intra-S* discourse relations, the class of explicit connectives has been extended to include subordinators — prepositions such as *because of, by, despite, for, from, in, instead of, regardless of, with* and *without*. However, they are only annotated when they take clausal complements. So *despite* would be annotated as an explicit connective in (6), but left unannotated in (7).

- (6) Digital Equipment gained 1 3/8 to 89 3/4 despite reporting earnings for the September quarter that were on the low end of expectations.

- (7) Despite recent declines in yields, investors continue to pour cash into money funds.

Other subordinators included as explicit connectives are *in order, no matter, so much as* and *notwithstanding*.

As with many of the subordinating conjunctions and some discourse adverbials, subordinators appear in the corpus in a modified form, as in *partly because of, not because of, merely by, only by, especially with, even with*, etc. (cf. Appendix B). Except for *mostly* (as in *mostly by*), all these modifiers (including *not*, as in *not because* and *at least not when*) also appear with subordinating conjunctions.

## 2.6 Spans that can be analyzed as one connective or two

Four sequences (*but then, only to, or otherwise* and *but also*) have been found analyzable as either two separate connectives or a single connective, with the single connective having a different sense distribution than either of the two separate connectives. As such, this is a new disambiguation task for discourse parsers.

This wasn't the case in the PDTB-2: All *but then* spans were analyzed as comprising two separate connectives: with *but* conveying CONTRAST or CONCESSION and *then* usually conveying TEMPORAL.PRECEDENCE, as in

- (8) Small businesses say a recent trend is like a dream come true: more-affordable rates for employee-health insurance, initially at least. But then they wake up to a nightmare. [wsj\_0518]

Our review of CONTRAST and CONCESSION tokens for the PDTB-3 (Section 4.2.3) showed that some of these tokens should be analyzed as a single connective expressing CONCESSION, as in

(9) *To many, it was a ceremony more befitting a king than a rural judge seated in the isolated foothills of the southern Allegheny Mountains. But then **Judge O’Kicki often behaved like a man who would be king – and, some say, an arrogant and abusive one.** [wsj\_0267]*

(10) *At Jefferies & Co., J. Francis Palamara didn’t reach the office until 5:30 a.m., but then **he had been up most of the night at home.** [wsj\_2205]*

Similarly, the sequence *only to* was also found to have two distinct analyses. In one, *only* is simply a modifier, conveying that Arg2 is the only thing in the given relation with Arg1, as in

(11) *Tax-exempt airport and street-corner solicitations were intended **only to provide start-up funds.*** [wsj\_0282]

In the other, *only to* is a single connective conveying that Arg2 is a surprising, unexpected situation revealed by Arg1, as in

(12) *Two former secretaries told the grand jury they were summoned to the judge’s chambers on separate occasions to take dictation, only to **find the judge in his bikini underwear.** [wsj\_0267]*

This is indicated by *only to* being labelled both CONCESSION.ARG2-AS-DENIER and PRECEDENCE.

The sequence *or otherwise* can also be analyzed either as two distinct connectives, with *otherwise* expressing a negative-condition ARG1-AS-NEGCOND, as in Example 13 (paraphrasable as “if you don’t stay in the center of the path, you might trip and fall.”), or as a single connective, with *or otherwise* conveying that the disjuncts are mutually exclusive **and** that their union covers the full set, as in Example 14.

(13) *Walk down the center of the path, or otherwise, **you might trip and fall.***

(14) *A new Maryland law frees store owners of liability if a customer *trips* or otherwise **gets hurt** on the way to the restroom. [wsj\_1270]*

Finally, the sequence *but also* (or in some cases, just the word *but*), sometimes appears as part of the paired connective *not only ... but also* (cf. Ex. 15), where (as often happens) elements of the pair may be missing (cf. Ex. 16-17), usually sense-annotated as CONJUNCTION.

(15) *Market participants say *investors are* not only *licking their wounds following the turbulence last week,* but **they have also been made nervous by two events in West Germany.** [wsj\_1187]*

(16) *Paul is no stranger to lavish spending,” the magazine said, noting that he *doesn’t stop at paint on canvas* but also **spends big on art you can eat.** [wsj\_2113]*

(17) *Not only do the actors stand outside their characters and make it clear they are at odds with them, but **they often literally stand on their heads.** [wsj\_1163]*

However, the span may also simply be analyzed as two distinct connectives — *but*, signalling CONTRAST or CONCESSION.ARG2-AS-DENIER, and *also*, signalling CONJUNCTION. In this case, each connective would be separately annotated, with the two then being **linked** (cf. Section 8.5).

(18) a. Editorials in the Greenville newspaper *allowed that Mrs. Yeargin was wrong,* but **also said the case showed how testing was being overused.** [wsj\_0044]

b. Editorials in the Greenville newspaper *allowed that Mrs. Yeargin was wrong,* **but** also **said the case showed how testing was being overused.** [wsj\_0044]

## 2.7 New Relation Types

### 2.7.1 Hypophora

We have added a new type of coherence relation for Question-Answer pairs, which we have labelled HYPOPHORA. In HYPOPHORA relations, one argument (commonly Arg1) expresses a question and the other argument (commonly Arg2) provides an answer. As with Entity Relations, no explicit or implicit connective is identified and annotated: The only elements of the relation are Arg1 and Arg2.

(19) *If not now, when?* “**When the fruit is ripe, it falls from the tree by itself,**” he says.” [wsj\_0300]

(20) *Of all the ethnic tensions in America, which is the most troublesome right now?* **A good bet would be the tension between blacks and Jews in New York City.** [wsj\_2369]

RelType HYPOPHORA should be chosen when the question seeks information, and the response aims to fulfill that need by addressing it explicitly, as in the previous examples, or implicitly, as in

(21) *But can Mr. Hahn carry it off?* **In this instance, industry observers say, he is entering uncharted waters.** [wsj\_0100]

(22) *So can a magazine survive by downright thumbing its nose at major advertisers?* **Garbage magazine, billed as ”The Practical Journal for the Environment,” is about to find out.** [wsj\_0062]

The response can also indicate that the information need cannot be fulfilled, as in

(23) *With all this, can stock prices hold their own?* ”**The question is unanswerable at this point**” she says. [wsj\_0681]

The relation type HYPOPHORA does not apply when the subsequent text relates to a question in other ways – for example, in the case of rhetorical questions that are posed for dramatic effect or to make an assertion, rather than to elicit an answer:

(24) *Remember Pinnochio?* (Implicit=similarly (Similarity)) **Consider Jim Courtier.** [wsj\_0041]

If the subsequent text is not an answer (direct or indirect) or a denial that an answer is possible, another relation should also be annotated:

(25) *Since chalk first touched slate, schoolchildren have wanted to know: What’s on the test?*  
(Implicit=however (Arg2-as-denier)) **These days, students can often find the answer in test-coaching workbooks and worksheets their teachers give them in the weeks prior to taking standardized achievement tests.** [wsj\_0045]

### 2.7.2 AltLexC

The new Relation Type – AltLexC – has been introduced to allow us to record where, within a sentence, a lexico-syntactic construction has been recognized as signalling a discourse relation. Since researchers may be interested in analyzing these constructions further, we have marked them with the relation type ALTLXEC for easy identification. Tokens of this type have all the same fields as an AltLex. Tokens of the following constructions have been marked as ALTLXEC in the PDTB-3:

- *so* ⟨Adj/Adv⟩ [*that*], expressing RESULT (Ex. 26)
- auxiliary inversion, expressing CONDITION.ARG2-AS-COND (Ex. 27)
- *such* ⟨NP⟩ [*that*], expressing RESULT (Ex. 28)
- *too* ⟨Adv⟩ [*that*], expressing NEGATIVE-RESULT (Ex. 29)
- *so* [*too*] ⟨aux⟩ ⟨NP⟩, expressing SIMILARITY (Ex. 30)
- predicate inversion, expressing CONCESSION.ARG1-AS-DENIER (Ex. 31)
- *as* ⟨aux⟩ ⟨NP⟩ and ⟨be⟩ ⟨VP/AdjP⟩, expressing SIMILARITY (Ex. 32–33)

(26) *The fit is so good, we see this as a time of opportunity.* [wsj\_0317]

(27) *... but would have climbed 0.6%, had it not been for the storm* [wsj\_0573]

(28) *A Ford takeover of Jaguar would “have such implications for the balance of power in the 1990s that General Motors can’t afford to step aside.* [wsj\_1225]

(29) *Things have gone too far for the government to stop them now* [wsj\_2454]

(30) *All independent media activity is now illegal but so is the manufacture of perfume, cosmetics, household chemicals and sand candles.* [wsj\_0439]

(31) *Crude as they were, these early PCs triggered explosive product development in desktop models for the home and office.* [wsj\_0022]

(32) *Bell Atlantic posted a strong earnings gain for the third quarter, as did Southern New England Telecommunications.* [wsj\_1728]

(33) *Mr. Wertheimer said the Senate Ethics Committee should hire a special outside counsel to conduct an investigation, as was done in the case of former House Speaker James Wright.* [wsj\_2446]

The distribution of AltLexC constructions in the corpus, along with their senses, is given in Appendix E. Researchers might discover other lexico-syntactic constructions in the corpus that should be noted as indicators of discourse relations – for example, ones discussed in (Dunietz et al., 2017). This is a rich area for further work.

### 3 What’s different in the PDTB-3?

Differences range from differences in the conventions used in annotation, to differences in the annotation of particular tokens, reflecting *inter alia* (1) changes and/or additions to the sense hierarchy; (2) different criteria for choosing one sense label over another; and (3) rigorous attention to enforcing as much consistency as possible in the annotation of discourse relations.

Of the ~53600 annotated tokens in the PDTB-3, ~57% are unchanged from their annotation in the PDTB-2. (Of these, ~42% are explicit tokens, ~39% implicit tokens, ~17% EntRel tokens, ~2% AltLex

tokens and ~1% NoRel tokens.) Another ~19% of the tokens in the PDTB-3 arise from PDTB-2 tokens that have had their annotation changed from what it was in the PDTB-2. (Of these, ~56% are explicit tokens that have been changed in some way, while ~40% are implicit tokens that have been changed. The other ~4% comprise changes to Entity Relations, AltLex relations or NoRel tokens.) The PDTB-3 release also lists 116 tokens that have been deleted from the PDTB-2. Most correspond to tokens whose Relation Type has changed (e.g., Implicit changed to Explicit or AltLex or AltLexC), so the version with the previous type appears to be deleted. (Information on the *provenance* of each token can be found in field 32 of the pipe-delimited file format described in Section 8.1.)

The rest of this section provides brief descriptions of what is different in the PDTB-3, which is then described in greater detail later on.

### 3.1 Labelling Arguments and their spans

In the PDTB-2, arguments were labelled according to the following convention.

- For spans linked by an **explicit** discourse connective, Arg2 was the argument to which the connective was attached syntactically, and the other was Arg1. This allowed consistent labelling of the arguments to subordinating conjunctions, independent of the order in which the arguments appeared. The same was true for coordinating conjunctions, whose argument order is always the same, and for discourse adverbials, whose Arg1 always precedes the adverbial, even when Arg1 is embedded in Arg2.
- For **implicit** discourse relations, Arg1 was always the first (lefthand) span and Arg2, the adjacent (righthand) span.

However, applying this convention in annotating *Intra-S* discourse relations can produce inconsistent labelling because explicit connectives can vary in where they attach within a sentence, as can the syntactic constructions that serve as evidence for discourse relations (cf. Section 2.7.2) .

The first problem can be illustrated with paired connectives like *not only . . . but also*. Here, both members of the pair may be present (Ex. 34), or just one or the other (Ex. 35 and Ex. 36):

(34) Japan not only outstrips the U.S. in investment flows but also outranks it in trade with most Southeast Asian countries . . . [wsj\_0043]

(35) The hacker was pawing over the Berkeley files but also using Berkeley and other easily accessible computers as stepping stones . . . [wsj\_0257]

(36) Not only did Mr. Ortega's comments come in the midst of what was intended as a showcase for the region, it came as Nicaragua is under special international scrutiny . . . [wsj\_0655]

A labelling convention that requires **Arg2** to be the argument to which the explicit connective attaches will choose a different argument for **Arg2** in Ex. 35 than in Ex. 36, and an arbitrary argument in the case of Ex 34, when semantically, the lefthand argument plays the same role in all three cases, as does the righthand argument.

The second problem can be illustrated with preposed auxiliaries, which signal a *CONDITIONAL* relation between the clause with the preposed auxiliary (as *antecedent*) and the other clause (as *consequent*). As with subordinating clauses, the two clauses can appear in either order:

- (37) Had the contest gone a full seven games, ABC could have reaped an extra \$10 million in ad sales . . . [wsj\_0443]
- (38) . . . they probably would have gotten away with it, had they not felt compelled to add Ms. Collins’s signature tune, “Amazing Grace,” . . . [wsj\_0207]

Since there is no explicit connective in either clause, if position is used to label *Arg1* and **Arg2**, the result can again be inconsistent.

To avoid inconsistency, while not requiring any change to existing labels in the PDTB-2, the argument-labelling convention has been changed as follows:

- The arguments to *Inter-S* discourse relations remain labelled by *position*: *Arg1* is first (lefthand) argument and **Arg2**, the second (righthand) argument.
- With *Intra-S* coordinating structures, the arguments are also labelled by *position*: *Arg1* is first argument and **Arg2**, the second one, independent of which argument(s) have attached coordinating conjunction(s).
- With *Intra-S* subordinating structures, *Arg1* and **Arg2** are determined syntactically. The subordinate structure is always labelled **Arg2**, and the structure to which it is subordinate is labelled *Arg1*.

This way of labelling subordinating structures allows for preposed and postposed clauses to have different sense distributions. This is most striking with CONCESSION — where *Arg1* can be taken to deny the expectation raised by **Arg2**, or vice versa. It can be seen in the sense distribution of *although*, *though* and *even though*, as well as with *when*, *if*, and *while*. For example, for preposed *although* clauses, about 75% of the ~200 tokens have been sense-annotated as ARG1-AS-DENIER, while with postposed *although* clauses, over 95% of the ~110 tokens have been sense-annotated as ARG2-AS-DENIER. While both cases show the second (right-hand) clause being taken to deny or cancel the expectation raised by the first (left-hand) clause, when *although* is preposed, *Arg1* is on the right (ARG1-AS-DENIER), while when *although* is postposed, *Arg2* is on the right (ARG2-AS-DENIER).

Finally, with respect to argument spans, the PDTB-2 required that each argument span subsume *at least one full clause*. While this continues to be the guideline for non-coordinating constructions, for *Intra-S* coordinating conjunctions, annotators are meant to annotate the conjuncts, so that the arguments are parallel. For conjoined VPs, therefore, the spans are *not* a whole clause, but just VPs: So in Ex. 39, *Arg1* subsumes only *joined the firm in 1963*, and not the subject *he*.

- (39) He *joined the firm in 1963* and **bought it from the owners the next year**. [wsj\_0305]

### 3.2 Identifying AltLex

ALTERNATIVE LEXICALIZATIONS (or ALTLEX) were only annotated in the PDTB-2 in the absence of an explicit connective, if annotators inferred a relation between sentences but felt that the insertion of an implicit connective would be redundant. In that case, they were encouraged to identify as an *AltLex*, the non-connective expression in *Arg2* that they took as the source of the perceived redundancy. (AltLex tokens can be distinguished from Implicit tokens by having ALTLEX as their relation type (RelType)).

Annotating *Intra-S* discourse relations in the PDTB-3 has led to modifying the above convention in two ways — what is annotated as ALTLEX and where ALTLEX can be annotated.

With respect what is annotated as ALTLEX, AltLex expressions in the PDTB-2 required one part that conveyed the relation and one part that referred anaphorically or elliptically to Arg1, as in “after that” or “a likely reason for the disparity is”. To allow for AltLex expressions in *Intra-S* discourse relations, we have allowed expressions of **any** form or syntactic class, including adjectives and adjective-modifiers such as *additional*, *next*, *further*, and *earlier*. While these expressions continue to suggest the relation, the reference to Arg1 may be implicit. That is, while *next* implies next to something, that something may be implicit.

One consequence of this is that words such as *further* and *next*, that can appear as discourse adverbials, or as adverbials modifying verbs or adjectives, or as adjectives themselves, will be annotated as Explicit connectives when they are discourse adverbials and as AltLex phrases otherwise, as in the following token, where *further* modifies *estimated*.

- (40) *According to Upjohn’s estimates, only 50% to 60% of the 1,100 eligible employees will take advantage of the plan. Upjohn **further** estimated that about 50% of the employees who leave for early retirement may be replaced.* [wsj\_0184]

This is in contrast with the following token of *further*, which serves as a discourse adverbial and is annotated as an Explicit connective.

- (41) Stephen G. Jerritts, president and chief executive officer, said *customers weren’t willing to commit to an expensive NBI hardware systems because of the company’s financial troubles. **Further**, he said, argii[the company doesn’t have the capital needed to build the business over the next year or two].* [wsj\_0092]

With respect to where *AltLex* can be annotated, there are two points of difference with the PDTB-2. First, PDTB-3 annotators have been permitted to include material from both Arg1 and Arg2 in the AltLex expression, as in

- (42) *Marni Rice plays the maid with so much edge **as to steal her two scenes.*** [wsj\_1163]
- (43) *some of the proposals are so close **that non-financial issues such as timing may play a more important role*** [wsj\_0953]
- (44) *Things have gone too far **for the government to stop them now.*** [wsj\_2454]

More on this appears in the discussion of lexico-syntactic constructions as Alternative Lexicalizations (Section 2.7.2).

The second difference is that AltLex expressions can also be annotated when there is already an Explicit connective, as in Ex. 45 and Ex. 46, or with an already inferred relation (Implicit connective), as in Ex. 47.

- (45) a. *... nor can the defendant be compelled to take the stand as a witness, thus **forcing him to “take the Fifth”*** [wsj\_2377]
- b. *... nor can the defendant be compelled to take the stand as a witness, **thus** forcing him to “take the Fifth”* [wsj\_2377]

- (46) a. *Admittedly last season’s runaway hit, “Steel Magnolias,” helped a lot, but so did cost cutting and other measures insisted on by the board.* [wsj\_0819]
- b. *Admittedly last season’s runaway hit, “Steel Magnolias,” helped a lot, but so did cost cutting and other measures insisted on by the board.* [wsj\_0819]

In addition to the explicit connective (*thus* expressing RESULT in Ex. 45 and *but* expressing CONTRAST in Ex. 46), a separate AltLex token has been annotated for *forcing* (with sense=RESULT) in Ex. 45, and for *so did* (with sense=SIMILARITY) in Ex. 46.

- (47) a. *... the government should encourage home ownership, (Implicit=by means of) including issuing bonds that guarantee holders the right to purchase an apartment.*
- b. *... the government should encourage home ownership, including issuing bonds that guarantee holders the right to purchase an apartment.*

Here, in addition to the inferred relation of EXPANSION.MANNER.ARG2-AS-MANNER, a separate AltLex token has been annotated for *including*, with sense EXPANSION.INSTANTIATION.ARG2-AS-INSTANCE.

### 3.3 Attribution

The PDTB-2 annotation of attribution contained a range of features, including whether the content should be attributed to the writer (Wr) or to some other agent mentioned in the text (Ot). These features were not annotated along with attribution in the PDTB-3, although attribution spans themselves have still been annotated where appropriate.

### 3.4 Multiple Relations between Arguments

The PDTB-2 allowed for more than one sense relation to hold between a pair of arguments in any of three conditions: (1) the senses could be associated with an explicit discourse connective that concurrently conveyed more than one sense – e.g., *since* conveying both a temporal relation and causal (REASON) relation; (2) each sense could be associated with a different explicit connective with essentially the same arguments – e.g., *but then*, with *but* conveying one relation between the arguments and *then* expressing another; (3) annotators could infer multiple senses as holding concurrently between two spans, in the absence of an explicit connective.

The PDTB-3 allows for a fourth case: annotators can also record a distinct implicit relation holding in the context of an explicit connective or AltLex that doesn’t itself convey that relation. The implicit relation is annotated on a separate token. More information about multiple relations can be found in Section 8.5.

### 3.5 Annotating Contrast vs Concession

The PDTB-3 reflects a more systematic convention for distinguishing between the Comparison relations, Contrast and Concession. This is discussed in Section 4.2.3.

### 3.6 Contents of the Conn field

In the underlying representation of the corpus, a relation token is represented in the pipe-delimited format described in Section 8.1. As in the PDTB-2, the first field indicates the relation type (e.g., Explicit, Implicit, AltLex, EntRel, etc.). In the PDTB-3 however, the contents of the second field depends on the relation type. For Explicit relations, the field indicates the span of the explicit connective and its modifiers (if any). For AltLex relations, the field indicates the span of the alternative lexicalization that signals the sense of the relation. This material can be located in Arg2 or Arg1 or both (see Section 3.2). For the new relation type AltLexC (see Section 2.7.2), the field points to the same span as Arg2 – that is, the lexico-syntactic construction that (unambiguously) conveys the sense of the relation. Users should filter the contents of the field accordingly, in order to get at the data they want.

### 3.7 Adverbials removed as Explicit Discourse Connectives

While many connectives (both subordinators and discourse adverbials) have been added to the set of Explicit Connectives, a few (*in turn*, *overall*, and *in general*) have been removed, and PDTB-2 relations containing them have been re-annotated.

In the case of *in turn*, while clearly meaning-bearing, it did not appear to be the source of the many different discourse relations posited between a clause containing *in turn* and some previous clause. That is, while *in turn* (1) can combine with an explicit conjunction (e.g., *but in turn*, *so in turn*, or *in turn*, or AltLex, as in the example *which is, in turn, why ...*, or (2) appear on its own, the senses associated with such tokens in the PDTB-2 (e.g., PRECEDENCE, RESULT, CONTRAST, CONJUNCTION or some combination thereof) do not seem to come from *in turn*, but are inferred from the argument spans and thus should not be associated with *in turn* itself. This can be seen by dropping *in turn*, and noticing that the same sense continues to hold. But because neither the PDTB-2 nor the PDTB-3 sense hierarchy contains any specific senses that might be conveyed by *in turn* (such as reciprocity or reciprocal causality), *in turn* has been removed from the set of explicit discourse connective in the PDTB-3, and tokens containing it have been systematically and consistently annotated as implicit relations.

Tokens previously annotated with *overall* as their explicit connective were reviewed and re-annotated as having a more informative, consistently annotated implicit connective. For example

(48) *The cross-border loan portfolio reflected “adjustment problems and episodic payment patterns,” the bank said no interest payments from Argentina in the nine months and none from Brazil in the third quarter, while Venezuela brought itself “substantially current”.* **Overall, the portfolio narrowed its quarterly loss to \$70 million from \$80 million a year earlier.** [wsj\_2040, ARG1-AS-DETAIL]

(49) *New telephone lines posted healthy growth.* **Overall they increased 2.8% to 12.1 million.** [wsj\_1651, ARG2-AS-DETAIL]

Finally, the phrase *in general* is now taken as an Alternative Lexicalization, and only in those cases where it plays a role in relating two clauses has it been annotated. Reviewing tokens of *in general* showed that the adverbial rarely related a specific description with a general one, as if it were the reverse of *in particular*, signalling the relation ARG1-AS-DETAIL. As such, *in general* was dropped from the set of explicit connectives as well. The set of terms annotated as Explicit Connectives appears in Appendix A.

## 4 PDTB-3 Senses

### 4.1 Differences between the PDTB-2 and PDTB-3 sense hierarchy

The PDTB-3 relation hierarchy simplifies and extends the PDTB-2 relation hierarchy (Table 1). Simplifications involve restricting Level-3 relations to differences in directionality and eliminating rare and/or difficult-to-annotate senses (Section 4.1.1). Additions permit annotating some of the new intra-sentential relations (Section 4.1.2).

#### 4.1.1 Simplifying the relation hierarchy

Although the hierarchy retains the same four Level-1 relations, relations at Level-3 now only encode *directionality* and so now only appear with asymmetric Level-2 relations.<sup>1</sup> Those Level-3 relations in the PDTB-2 that did not have a direction (e.g., EQUIVALENCE) were either moved to Level-2 or eliminated due to their rarity or the difficulty they posed for annotators (e.g., the Level-3 relations under CONTRAST).

With respect to directionality, some relations whose arguments occur in a single order inter-sententially in the PDTB-2, turn out to occur in either order intra-sententially. This means that they need two Level-3 versions. For example, the argument conveying the condition in CONDITION relations can be either **Arg2** (as was the case throughout the PDTB-2) or *Arg1* as in Ex. 50, while the argument conveying the CHOSEN ALTERNATIVE (now called SUBSTITUTE) in SUBSTITUTION relations can be either **Arg2** (as was the case throughout the PDTB-2) or *Arg1*, as in Ex. 51. With EXCEPTION, it hadn't been noticed that in some of the tokens so annotated, the exception appeared in **Arg2**, while in others, the exception appeared in *Arg1*. The difference is now supported with a distinct Level-3 type in each direction (Exs. 52–53).

- (50) ARG1-AS-COND: *Call Jim Wright's office in downtown Fort Worth, Texas, these days and **the receptionist still answers the phone,** "Speaker Wright's office.* [wsj\_0909]
- (51) ARG1-AS-SUBST: *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* [wsj\_0443]
- (52) ARG1-AS-EXCPT: *Twenty-five years ago the poet Richard Wilbur modernized this 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.*** [wsj\_1936]
- (53) ARG2-AS-EXCPT: *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.*** [wsj\_1103]

Level-2 pragmatic relations have been removed from the PDTB-2 and replaced with relations whose label indicates whether *implicit* belief (epistemic knowledge) or a *speech act* is associated with arguments (cf. Table 1, which shows the relations for which evidence has been found for a belief or speech act version). For example, Ex. 54 shows an implicit CAUSE.RESULT relation in which the result **Arg2** argument is the (speaker's/writer's) *belief* that the deadline could be extended. Its sense is therefore labelled CONTINGENCY.CAUSE+BELIEF.RESULT+BELIEF. Similarly, Ex. 55 shows a CONCESSION.ARG2-AS-DENIER relation, in which what is denied (or cancelled) is the speech act associated with **Arg2**. Its sense is therefore labelled COMPARISON.CONCESSION+SPEECHACT.ARG2-AS-DENIER+SPEECHACT.

<sup>1</sup>A sense relation  $\mathfrak{R}$  is *symmetric* iff  $\mathfrak{R}(Arg1, \mathbf{Arg2})$  and  $\mathfrak{R}(\mathbf{Arg2}, Arg1)$  are semantically equivalent. If a relation is not symmetric, it is *asymmetric*.

Level-1	Level-2	Level-3
TEMPORAL	SYNCHRONOUS	–
	ASYNCHRONOUS	PRECEDENCE
		SUCCESSION
CONTINGENCY	CAUSE	REASON
		RESULT
		NEGRESULT
	CAUSE+BELIEF	REASON+BELIEF
		RESULT+BELIEF
	CAUSE+SPEECHACT	REASON+SPEECHACT
		RESULT+SPEECHACT
	CONDITION	ARG1-AS-COND
		ARG2-AS-COND
	CONDITION+SPEECHACT	–
	NEGATIVE-CONDITION	ARG1-AS-NEGCOND
		ARG2-AS-NEGCOND
	NEGATIVE-CONDITION+SPEECHACT	–
	PURPOSE	ARG1-AS-GOAL
ARG2-AS-GOAL		
COMPARISON	CONCESSION	ARG1-AS-DENIER
		ARG2-AS-DENIER
	CONCESSION+SPEECHACT	ARG2-AS-DENIER+SPEECHACT
	CONTRAST	–
SIMILARITY	–	
EXPANSION	CONJUNCTION	–
	DISJUNCTION	–
	EQUIVALENCE	–
	EXCEPTION	ARG1-AS-EXCPT
		ARG2-AS-EXCPT
	INSTANTIATION	ARG1-AS-INSTANCE
		ARG2-AS-INSTANCE
	LEVEL-OF-DETAIL	ARG1-AS-DETAIL
		ARG2-AS-DETAIL
	MANNER	ARG1-AS-MANNER
		ARG2-AS-MANNER
SUBSTITUTION	ARG1-AS-SUBST	
	ARG2-AS-SUBST	

Table 1: PDTB-3 Sense Hierarchy. The leftmost column contains the Level-1 senses and the middle column, the Level-2 senses. For asymmetric relations, Level-3 senses are located in the rightmost column.

(54) That deadline *has been extended once* and Implicit=so **could be extended again.** [wsj\_2032]

(55) He spends his days *sketching passers-by*, or **trying to.** [wsj\_0039]

The LIST relation has been eliminated from the PDTB-2 hierarchy, since it was not in practice distinguishable from CONJUNCTION. And the names of two asymmetric PDTB-2 relations have been changed to bring out commonalities: RESTATEMENT has been renamed LEVEL-OF-DETAIL, with its SPECIFICATION and GENERALIZATION subtypes now just taken to be directional variants — ARG2-AS-DETAIL and ARG1-AS-DETAIL, respectively; and the sub-types of CONCESSION, opaquely called CONTRA-EXPECTATION and EXPECTATION, have been renamed to reflect simply a difference in directionality — ARG1-AS-DENIER and ARG2-AS-DENIER.

#### 4.1.2 Augmenting the relation hierarchy

Additional senses were found to be needed for annotating *Intra-S* discourse relations. These include the asymmetric MANNER relation under EXPANSION, and a separate ARG1-AS-INSTANCE relation under INSTANTIATION. Under CONTINGENCY, additional senses were found needed for the asymmetric Level-2 relations of NEGATIVE\_CONDITION and PURPOSE (both having Arg1 and Arg2 directions at Level-3) and a Level-3 relation of NEGATIVE\_RESULT under CAUSE. (No evidence has been found for a NEGATIVE\_REASON sense.) The new symmetric Level-2 relation of SIMILARITY was added under COMPARISON because of its obvious omission from the PDTB-2 as the complement of the symmetric relation CONTRAST.

Note that, as well as being used to annotate new tokens, all the existing tokens in the PDTB-2 have been mapped to senses in the revised relation hierarchy. While most of the mapping is simply 1:1, where it is not, manual review has been required to ensure both agreement and consistency (cf. Section 6.2).

## 4.2 Sense Classification

Here we provide definitions of the PDTB-3 sense labels shown in Table 1.

### 4.2.1 Temporal Relations

The tag TEMPORAL is used when the situations described in the arguments are intended to be related temporally.

**Temporal.Synchronous** This tag is used when there is some degree of temporal overlap between the events described by the arguments. All forms of overlap are included.

(56) *Knowing a tasty – and free – meal* when **they eat one**, the executives gave the chefs a standing ovation. [wsj\_0010]

(57) The company is operating under Chapter 11 of the federal Bankruptcy Code, *giving it court protection from creditors' lawsuits* while **it attempts to work out a plan to pay its debts.** [wsj\_0031]

(58) *Then, in late-afternoon trading, hundred-thousand-share buy orders for UAL hit the market, including a 200,000-share order through Bear Stearns that seemed to spark UAL's late price surge.* Almost simultaneously, **PaineWebber began a very visible buy program for dozens of stocks.** [wsj\_1208]

(59) *But as the fright began to spread through the S&P pit, the big brokerage firms came in and bought futures aggressively.* [wsj\_1208]

(60) *The parishioners of St. Michael and All Angels stop to chat at the church door, as members here always have. (Implicit=while) **In the tower, five men and women pull rhythmically on ropes attached to the same five bells that first sounded here in 1614.*** [wsj\_0089]

**Temporal.Asynchronous** This tag is used when one event is described as preceding the other.

The label TEMPORAL.ASYNCHRONOUS.PRECEDENCE is used when the event described by Arg1 precedes that described by Arg2 (i.e., Arg1 << Arg2).

(61) *A buffet breakfast was held in the museum, where food and drinks are banned to everyday visitors. Then, in the guests' honor, the speedway hauled out four drivers, crews and even the official Indianapolis 500 announcer for a 10-lap exhibition race.* [wsj\_0010]

(62) *Back downtown, the execs squeezed in a few meetings at the hotel before boarding the buses again.* [wsj\_0010]

(63) *Output will be gradually increased until it reaches about 11,000 barrels a day.* [wsj\_0024]

(64) *The Artist has his routine. He spends his days sketching passers-by, or trying to. (Implicit=then) **At night he returns to the condemned building he calls home.*** [wsj\_0039]

The label TEMPORAL.ASYNCHRONOUS.SUCCESION is used when the event described by Arg2 precedes that described by Arg1 (i.e., Arg2 << Arg1).

(65) *William Gates and Paul Allen in 1975 developed an early language-housekeeper system for PCs, and Gates became an industry billionaire six years after **IBM adapted one of these versions in 1981.*** [wsj\_0022]

(66) *John D. Carney, 45, was named to succeed Mr. Hatch as president of Eastern Edison. Previously he was vice president of Eastern Edison.* [wsj\_0019]

(67) *(Implicit=when) **Pressed on the matter**, he is more specific.* [wsj\_0300]

#### 4.2.2 Contingency Relations

The tag CONTINGENCY is used when the situation described by one argument provides the reason, explanation or justification for the situation described by the other.

**Contingency.Cause** This tag is used when the situations described in Arg1 and Arg2 are causally influenced but are not in a conditional relation.

The label CONTINGENCY.CAUSE.REASON is used when Arg2 gives the reason, explanation or justification, while Arg1 gives its effect.

(68) *Runways at San Francisco weren't damaged, but traffic was being limited yesterday to 27 arrivals and 27 departures an hour – down from 33 to 45 an hour normally – mainly because **the noise level in the control tower was overwhelming without the windows**, an FAA spokeswoman said. [wsj\_1803]*

- (69) *But service on the line is expected to resume by noon today.* (Implicit=since) “**We had no serious damage on the railroad,**” said a Southern Pacific spokesman. [wsj\_1803]
- (70) *By 11:59 p.m. tonight, President Bush must order \$16 billion of automatic, across-the-board cuts in government spending to comply with the Gramm-Rudman budget law.* The cuts are necessary because Congress and the administration have failed to reach agreement on a deficit-cutting bill. [wsj\_2384]

The label CONTINGENCY.CAUSE.RESULT is used when *Arg1* gives the reason, explanation or justification, while **Arg2** gives its effect.

- (71) The bill would then declare *that the debt is equity and therefore isn’t deductible.* [wsj\_1822]
- (72) *Now, though, enormous costs for earthquake relief will pile on top of outstanding costs for hurricane relief.* “That obviously means that we won’t have enough for all of the emergencies that are now facing us, . . .,” Mr. Fitzwater said. [wsj\_1824]
- (73) “*We are going to explode lower,*” says the flamboyant market seer, . . . (Implicit=so) **Anyone telling you to buy stocks in this market is technically irresponsible.** [wsj\_0359]

Finally, the label CONTINGENCY.CAUSE.NEGRESULT is used when *Arg1* gives the reason, explanation or justification that prevents the effect mentioned in **Arg2**. It was specifically introduced for the lexico-syntactic construction *too X to Y* (cf. Section 2.7.2)

- (74) . . . *that Banco Exterior may have waited too long to diversify from its traditional export-related activities.* [wsj\_0616]
- (75) *Friday’s stock market sell-off came too late for many investors to act.* [wsj\_2306]

We have found no evidence in the corpus for the complementary relation CONTINGENCY.CAUSE.NEGREASON, so have not introduced it into the sense hierarchy.

**Contingency.Cause+Belief** This tag is used when evidence is provided to cause the hearer to believe a claim. The belief is implicit.

The label CONTINGENCY.CAUSE+BELIEF.REASON is used when *Arg2* gives the evidence justifying the claim given in *Arg1*.

- (76) With this sort of sentiment common, *it’s natural for investors to seek out “defensive” investment.* [wsj\_0359]
- (77) *The nations of southern Africa know a lot about managing elephants;* (Implicit=as) **their herds are thriving.** [wsj\_2047]
- (78) *And until last Friday, it seemed those efforts were starting to pay off.* (Implicit=because) “**Some of those folks were coming back,**” says Leslie Quick Jr., chairman, of discount brokers Quick & Reilly Group Inc. [wsj\_1866]

The label CONTINGENCY.CAUSE+BELIEF.RESULT is used when Arg1 gives the evidence justifying the claim given in Arg2.

(79) *Kellogg suspended work on a \$1 billion cereal plant, indicating a pessimistic outlook by the cereal maker, which has been losing market share.* [wsj\_0675]

(80) *There were no {sell} lists and the calendar is lightening up a bit. (Implicit=so) There's light at the end of the tunnel for municipalities.* [wsj\_0351]

**Contingency.Cause+SpeechAct** This tag is used when a reason is provided for the speaker uttering a speech act. The speech act is implicit.

The label CONTINGENCY.CAUSE+SPEECHACT.REASON is used when Arg2 is the reason for the speaker uttering the speech act in Arg1.

(81) *"Maybe I'm a little stuffy, but I wouldn't sell them," sniffs Bob Machon, owner of Papa's Sports Cards in Menlo Park, California.* [wsj\_1560]

The label CONTINGENCY.CAUSE+SPEECHACT.RESULT is used when Arg1 is the reason for the speaker uttering the speech act in Arg2.

(82) *I've seen a great many moviestar film portraits, and this one is outstanding.* [wsj\_0976]

(83) *Surviving scandal has become a rite of political passage at a time when a glut of scandal has blunted this town's sensibility. (implicit=so) Let the president demand strict new ethics rules.* [wsj\_0909]

**Contingency.Purpose** This tag is used when one argument presents an action that an AGENT undertakes with the purpose of the GOAL conveyed by the other argument being achieved. Usually (but not always), the agent undertaking the action is the same agent aiming to achieve the goal.

The label CONTINGENCY.PURPOSE.ARG1-AS-GOAL is used when Arg1 describes the goal and Arg2 describes the action undertaken to achieve it.

(84) *There are the strict monetarists, who believe that floating exchange rates free an economy *to stabilize its price level* by stabilizing the monetary aggregate.* [wsj\_0553]

(85) *She ordered *the foyer done in a different plaid planting*, and (Implicit=for that purpose) made the landscape architects study a book on tartans.* [wsj\_0984]

The label CONTINGENCY.PURPOSE.ARG2-AS-GOAL is used when Arg2 describes the goal and Arg1 describes the action undertaken to achieve it.

(86) *Skilled ringers use their wrists to advance or retard the next swing, so that one bell can swap places with another in the following change.* [wsj\_0089]

(87) *In September, the company said it was seeking offers for its five radio stations in order to concentrate on its programming business.* [wsj\_0115]

(88) *I think *they should close down the futures exchange* and then we could get back to investing.* (wsj\_0742)  
(In this token from wsj\_0742, the agent of Arg1 is not the same as the agent whose goal is to get back to investing.)

**Contingency.Condition** This tag is used when one argument presents a situation as unrealized (the antecedent), which (when realized) would lead to the situation described by the other arg (the consequent). There are distinct senses for interpreting the arguments in terms of their semantics or the speech acts they convey. The default is their semantics.

The label CONTINGENCY.CONDITION.ARG1-AS-COND is used when Arg1 describes the antecedent and Arg2, the consequent.

- (89) *Call Jim Wright's office in downtown Fort Worth, Texas, these days and the receptionist still answers the phone, "Speaker Wright's office."* (wsj\_0909)
- (90) ...but he's still yelling: "*I find myself explaining anything to Teddy Kennedy, (implicit=then) you'll be chasing stolen cars in Anchorage*" (wsj\_1397)

While CONTINGENCY.CONDITION.ARG1-AS-COND should be considered when an imperative is *and*-conjoined with a declarative, as in Ex. 89, it is not always appropriate. In Ex. 91, there is no sense of CONDITION and REASON is more appropriate.

- (91) *Before your next California-bashing editorial, please spend more time out here witnessing the situation – (Implicit=because) it just may change your view.* [wsj\_0576]

The label CONTINGENCY.CONDITION.ARG2-AS-COND is used when Arg2 describes the antecedent and Arg1, the consequent.

- (92) But some bond market analysts said *that could quickly change if property casualty insurance companies scramble to sell portions of their municipal portfolios to raise cash to pay damage claims.* [wsj\_1874]
- (93) Trade and Industry Secretary Nicholas Ridley told the House of Commons yesterday *that he will relinquish the government's so-called golden share in the company] as long as Jaguar shareholders agree.* [wsj\_0224]
- (94) Had the contest gone a full seven games, *ABC could have reaped an extra \$10 million in ad sales on the seventh game alone*[wsj\_0443]
- (95) *At the same time, though, it must become more structured (implicit=so as) to better manage its growth.* [wsj\_0367]

**Contingency.Condition+SpeechAct** This tag is used when the consequent is an implicit speech act. So far, all cases of Condition+SA are Arg2-as-cond, so Arg1 is the implicit SA.

- (96) When *it comes to buying and selling shares, Westridge Capital Management Inc. takes a back seat to no one.* (wsj\_1600)
- (97) If *bringing the message is a crime, I'm guilty of it.* [wsj\_0349]

**Contingency.Negative-condition** This tag is used when one argument (the antecedent) describes a situation presented as unrealized, which if it doesn't occur, would lead to the situation described by the other argument (the consequent). There are distinct senses for interpreting the arguments in terms of semantics or speech acts, with the default being semantics.

The label CONTINGENCY.NEGATIVE-CONDITION.ARG1-AS-NEGCOND is used when Arg1 describes the antecedent and Arg2, the consequent.

- (98) The National Institutes of Health policy would require researchers to *cut financial ties with health-care businesses* or lose their government money. [wsj\_0975]
- (99) *This will prevent a slide in industrial production, which will otherwise cause new panic buying*. [wsj\_1646]

The label CONTINGENCY.NEGATIVE-CONDITION.ARG2-AS-NEGCOND is used when Arg2 describes the antecedent (aka condition) and Arg1, the consequent.

- (100) *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.*
- (101) Unless the Federal Reserve eases interest rates soon to stimulate the economy, *profits could remain disappointing*. (wsj\_0322)
- (102) Sandoz said *it expects a "substantial increase" in consolidated profit for the full year, barring major currency rate change*. [wsj\_2089]

**Contingency.Negative-condition+SpeechAct** This tag is used when the consequent is an implicit speech act. While none of the tokens in the PDTB-3 have been annotated with this sense, it is included in the hierarchy for completeness.

- (103) Unless you're on a diet, *there are some cookies in the cupboard*.

### 4.2.3 Comparison Relations

The tag COMPARISON is used when the discourse relation between two arguments highlights their differences or similarities, including differences between expected consequences and actual ones. The PDTB-3 sense hierarchy distinguishes three types of COMPARISON: CONTRAST, SIMILARITY and CONCESSION. SIMILARITY, the opposite of CONTRAST, had not been included in the PDTB-2 sense hierarchy. Of the other two, CONTRAST is meant to be used when at least **two** differences between Arg1 and Arg2 are highlighted, while CONCESSION is meant to be used when a causal relation expected on the basis of one argument is cancelled or denied by the situation described in the other. While these definitions appear quite different, in practice, annotators have had difficulty deciding which relation held between a pair of arguments. As such the following decision procedure was adopted when trying to decide between CONTRAST and CONCESSION:

1. Are at least two explicit differences highlighted between the arguments?
2. If no, select CONCESSION.

3. If yes, check whether a causal relation that is expected on the basis of one argument is denied by the other. (Test by paraphrasing with *although*.)
4. If yes, select CONCESSION.
5. If no, select, CONTRAST.

One consequence of this procedure is that, whenever CONCESSION can be taken as holding, it is annotated as such, even if CONTRAST also holds by definition. As this same procedure was also applied to tokens previously labelled CONTRAST or CONCESSION in the PDTB-2 (both Explicit and Implicit), the proportion of CONTRAST vs. CONCESSION tokens has changed significantly: Among explicit tokens of CONTRAST and CONCESSION in the PDTB-2, about 75% had been labeled CONTRAST and 25% CONCESSION. In contrast, the PDTB-3 has 22% labelled CONTRAST and 78% tokens labelled CONCESSION (either ARG1-AS-DENIER or ARG2-AS-DENIER). Among implicit tokens of CONTRAST and CONCESSION in the PDTB-3, ~40% are labelled CONTRAST and ~60% are labelled CONCESSION (either ARG1-AS-DENIER or ARG2-AS-DENIER).

**Comparison.Concession** This tag is used when an expected causal relation is cancelled or denied by the situation described in one of the arguments.

The label COMPARISON.CONCESSION.ARG1-AS-DENIER is used when it is Arg2 that raises an expectation of some consequence, and Arg1 that denies it.

- (104) The documents also said *that although the 64-year-old Mr. Cray has been working on the project for more than six years, the Cray-3 machine is at least another year away from a fully operational prototype.* [wsj\_0018]
- (105) *It's as if investors, the past few days, are betting that something is going to go wrong – even if they don't know what.* [wsj\_0359]
- (106) (Implicit=although) **Barely visible on Hong Kong's property scene in 1985, by last year Japan had become the top foreign investor.** [wsj\_0524]

The label COMPARISON.CONCESSION.ARG2-AS-DENIER is used when it is Arg1 that raises an expectation of some consequence, while it is Arg2 denies it.

- (107) *Last Friday, 96 stocks on the Big Board hit new 12-month lows. **But by Mr. Granville's count, 493 issues were within one point of such lows.*** [wsj\_0359]
- (108) *Several of those post-crash changes kicked in during Friday's one-hour collapse and worked as expected, even though they didn't prevent a stunning plunge.* [wsj\_2417]
- (109) *American Brands "just had a different approach," Mr. Wathen says. (Implicit=however) "Their approach didn't work."* [wsj\_0305]

**Comparison.Concession+SpeechAct** This tag is used when the speech act (SA) associated with one argument is cancelled or denied by the other argument or its SA. The only sub-type for which tokens have been identified is COMPARISON.CONCESSION+SPEECHACT.ARG2-AS-DENIER+SA, where it is the SA associated with Arg1 that is cancelled or denied by Arg2 or its associated SA. (N.B. The SA that is cancelled MUST be implicit.)

(110) *Congress closed this loophole last year, or **thought it did**.* (wsj\_1574)

This can be paraphrased as follows, with its implicit SAs made explicit: *While I say Congress closed this loophole last year, it might be more accurate to say Congress thought it closed this loophole last year.*

(111) *He lived in Peking, or **should I say Beijing, for 20 years**.*

Here, the Arg2 SA is explicit, although the Arg1 SA (the one that is denied) is (as required) implicit. It can be paraphrased as *While I say He lived in Peking, it might be more accurate to say he lived in Beijing.*

**Comparison.Contrast** As noted, CONTRAST is used when at least two differences between Arg1 and Arg2 are highlighted.

(112) *While **the earnings picture confuses**, observers say *the major forces expected to shape the industry in the coming year are clearer*.* [wsj\_2365]

(113) *After all, gold prices usually soar when inflation is high. **Utility stocks, on the other hand, thrive on disinflation** ...* [wsj\_0359]

(114) *Mr. Edelman said the decision "has nothing to do with Marty Ackerman." (implicit=on the contrary) **Mr. Ackerman contended that it was a direct response to his efforts to gain control of Datapoint**.* [wsj\_0333]

**Comparison.Similarity** This tag is used when one or more similarities between Arg1 and Arg2 are highlighted with respect to what each argument predicates as a whole or to some entities it mentions.

(115) *... that even after Monday's 10% decline, the Straits Times index is up 24% this year, so investors who bailed out generally did so profitably. Similarly, **Kuala Lumpur's composite index yesterday ended 27.5% above its 1988 close**.* [wsj\_2230]

(116) *Just as the 1980s bull market transformed the U.S. securities business, so too will **the more difficult environment of the 1990s**," says Christopher T. Mahoney, a Moody's vice president.* [wsj\_0128]

(117) *Builders get away with using sand (implicit=similarly) **and financiers junk** ...* [wsj\_1849]

#### 4.2.4 Expansion Relations

The label EXPANSION is used for relations that expand the discourse and move its narrative or exposition forward.

**Expansion.Conjunction** The tag CONJUNCTION is used when both arguments bear the same relation to some other situation evoked in the discourse. It indicates that the two arguments make the same contribution with respect to that situation, or contribute to it together. It differs from most other relations in that *the arguments don't directly relate to each other, but to this other situation*. For example, consider the conjuncts in Ex. 118, and then in the context of the previous sentence (Ex. 119).

(118) *I can adjust the amount of insurance I want against the amount going into investment;* (Implicit=CONJUNCTION)  
**I can pay more or less than the so-called target premium in a given year.** [wsj\_0041]

(119) *Hugely popular, it is far more flexible than straight whole life.* (Implicit=ARG2-AS-INSTANCE) **I can adjust the amount of insurance I want against the amount going into investment.** [wsj\_1574]

In Ex. 119, the first conjunct is linked to the situation in the previous sentence by an INSTANTIATION relation, as ARG2-AS-INSTANCE. From this can be inferred that the second conjunct of Ex. 118 is linked to this situation by an ARG2-AS-INSTANCE relation as well.

In the corpus, the *Wall Street Journal* style often mentions the relevant situation at the end of one paragraph, starting the next paragraph with the CONJUNCTION relation that elaborates it. If there is no explicit connective linking the start of one paragraph to the end of the previous one, the relation to that situation will not have been annotated in the PDTB-3. However, it will have been annotated in the cross-paragraph annotation supplement to the corpus described in (Prasad et al., 2017) and available at <https://github.com/pdtb-upenn/full-text>.

**Expansion.Disjunction** The tag EXPANSION.DISJUNCTION is used when the two arguments are presented as alternatives, with either one or both holding. As with CONJUNCTION, DISJUNCTION is used when both its arguments bear the same relation to some other situation evoked in the discourse, making a similar contribution with respect to that situation. While the arguments also relate to each other as alternatives (with one or both holding), they also both relate in the same way to this other situation. For example, consider the disjuncts in Ex. 120, and then in the context of the conditional clause (Ex. 121).

(120) *If we want to support students, we might adopt the idea used in other countries of offering more scholarships based on something called "scholarship," rather than on the government's idea of "service."* **Or we might provide a tax credit for working students.** [wsj\_2407]

(121) **If we want to support students,** *we might adopt the idea used in other countries of offering more scholarships based on something called "scholarship," rather than on the government's idea of "service."* [wsj\_2407]

In Ex. 121, the first disjunct is linked to the conditional clause in an ARG2-AS-COND relation. From this can be inferred that the second disjunct of Ex. 120 is linked to this same situation by an ARG2-AS-COND relation as well.

As with CONJUNCTION relations, a DISJUNCTION relation may be paragraph-initial, with the relevant situation mentioned in the previous paragraph. While the relation to this situation may not have been annotated in the PDTB-3, it will have been annotated in the cross-paragraph annotation supplement to the corpus described in (Prasad et al., 2017) and available at <https://github.com/pdtb-upenn/full-text>.

**Expansion.Equivalence** This tag is used when both arguments are taken to describe the same situation, but from different perspectives.

(122) 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.** [wsj\_0331]

(123) *But the battle is more than Justin bargained for.* (implicit=indeed) **"I had no idea I was getting in so deep,"** says Mr. Kaye, who founded Justin in 1982. [wsj\_2418]

**Expansion.Exception** This tag is used when one argument evokes a set of circumstances in which the described situation holds, and the other argument indicates one or more instances where it doesn't.

The label EXPANSION.EXCEPTION.ARG1-AS-EXCPT is used when Arg1 indicates the exception(s).

- (124) *Twenty-five years ago the poet Richard Wilbur modernized this 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.* [wsj\_0936]
- (125) *Some Japanese operations, such as securities-trading rooms, may be ahead of their American counterparts, he says, but (Implicit=otherwise) "basically, there's little analysis done on computers in Japan."* [wsj\_0445]

The label EXPANSION.EXCEPTION.ARG2-AS-EXCPT is used when Arg2 indicates the exception(s).

- (126) *Boston Co. officials declined to comment on Moodys action on the units 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.Instantiation** This tag is used when one argument describes a situation as holding in a set of circumstances, while the other argument describes one or more of those circumstances.

The label EXPANSION.INSTANTIATION.ARG1-AS-INSTANCE is used when Arg1 provides one or more instances of the circumstances described by Arg2.

- (127) *... and foreign companies such as Givaudan that deal with those industries are being hit the hardest. But in general, all foreign-trading companies are feeling the pinch.* [wsj\_1469]
- (128) *Then, as if to show that he could play fast as well, he offered the second movement from Saint-Saens's Sonata for Clarinet, ...* [wsj\_0207]
- (129) *Swiveling in his chair, Mr. Straszheim replies that the new outlook, though still weak, doesn't justify calling a recession right now. "It's all in this handout you don't want to look at. We could still have a recession" at some point. (Implicit=generally) One of Mr. Straszheim's recurring themes is that the state of the economy isn't a simple black or white.* [wsj\_0569]

The label EXPANSION.INSTANTIATION.ARG2-AS-INSTANCE is used when Arg2 provides one or more instances of the circumstances described by Arg1.

- (130) *The computers were crude by today's standards. Apple II owners, for example, had to use their television sets as screens and stored data on audiocassettes.* [wsj\_0022]
- (131) *And regional offices were "egregiously overstaffed," he claims. (Implicit=for example) One office had 19 people doing the work of three, ...* [wsj\_0305]

**Expansion.Level-of-Detail** This tag is used when both arguments describe the same situation, but in less or more detail.

The label EXPANSION.LEVEL-OF-DETAIL.ARG1-AS-DETAIL is used when Arg1 describes in more detail, the situation in Arg2. That is, Arg2 is more general.

- (132) *Many modern scriptwriters seem to be incapable of writing drama, or anything else, without foul-mouthed cursing. Sex and violence are routinely included even when they are irrelevant to the script, and high-tech special effects are continually substituted for good plot and character development. In short, we have a **movie and television industry that is either incapable or petrified of making a movie unless it carries a PG-13 or R rating.*** [wsj\_0911]

The label EXPANSION.LEVEL-OF-DETAIL.ARG2-AS-DETAIL is used when Arg2 describes in more detail, the situation in Arg1.

- (133) A Lorillard spokeswoman said, “*This is an old story. (Implicit = in fact) **Were talking about years ago before anyone heard of asbestos having any questionable properties.***” [wsj\_0003]
- (134) *An enormous turtle has succeeded where the government has failed: (Implicit = specifically) **He has made speaking Filipino respectable.*** [wsj\_0804]

**Expansion.Manner** This tag is used when the situation described by one argument presents the manner in which the situation described by other argument has happened or been done. MANNER answers “how” questions such as “How were the children playing?”. While MANNER may be the only relation that holds between two arguments, it is often the case that another sense (PURPOSE, RESULT or CONDITION) is taken to hold as well.

The label EXPANSION.MANNER.ARG1-AS-MANNER is used when it is Arg1 that describes the manner.

- (135) *McCaw is offering \$125 a share for 22 million LIN shares, thereby **challenging LIN’s proposal to spin off its television properties, pay shareholders a \$20-a-share special dividend and combine its cellular-telephone operations with BellSouth’s cellular business.*** [wsj\_2443]  
Arg1 answers the question *How did McCaw challenge LIN’s proposal?*

- (136) Long-debated proposals to *simplify the more than 150 civil penalties (Implicit=thereby) and make them fairer and easier to administer* are in the House tax bill. [wsj\_0293]  
Arg1 answers the question *How do the proposals make civil penalties fairer and easier to administer?*

The label EXPANSION.MANNER.ARG2-AS-MANNER is used when it is Arg2 that describes the manner.

- (137) Taking a cue from California, *more politicians will launch their campaigns by backing initiatives*, says David Magleby of Brigham Young University. [wsj\_0120]  
Arg2 answers the question *How are politicians launching their campaigns?*
- (138) In China, a great number of workers are engaged *in pulling out the male organs of rice plants (Implicit=by) using tweezers.* [wsj\_0209]  
Arg2 answers the question *How are workers pulling out the male organs of rice plants?*

Other senses are often taken to hold concurrently with either Arg1-as-manner or Arg2-as-manner. Among the explicit relations annotated ARG2-AS-MANNER, over half (52%) have been annotated with this sense alone, with the remaining 48% annotated both ARG2-AS-MANNER and another sense. Among implicit relations annotated either ARG1-AS-MANNER or ARG2-AS-MANNER, the figures are ~30% annotated with a MANNER sense alone, and ~70% annotated both MANNER and another sense. (The four examples used above to illustrate MANNER have been annotated with this sense alone.)

Among the most common senses taken to hold concurrently with a MANNER sense are ARG1-AS-GOAL, REASON, and ARG2-AS-COND holding with ARG2-AS-MANNER, and ARG2-AS-GOAL, RESULT and ARG2-AS-COND holding with ARG1-AS-MANNER. (The distribution of senses taken to hold concurrently can be found in Appendix G and Appendix I.)

(139) Mr. van Dover said *the AT&T team created the desired crystal changes* by [Arg1-as-goal, Arg2-as-manner] **bombarding superconductor samples with neutrons** ... [wsj\_0123]

(140) By [Reason, Arg2-as-manner] **lifting ethylene production**, *the expansion will also lower the company's raw material costs* [wsj\_2314]

(Paraphrase: *Not only is lifting ethylene production the manner in which raw material costs will be lowered, they will be lowered because ethylene production has been lifted.*)

(141) He said *that firms could get around the collar* by [Arg2-as-cond, Arg2-as-manner] **executing trades manually**.

(Paraphrase: *If firms executed trades manually, they could get around the collar, but executing trades manually is also the manner by which firms could do so.*)

(142) Some criminal lawyers speculated *that the IRS was sending the letters* [Arg2-as-goal, Arg1-as-manner] **to test the issue**. [wsj\_0049]

(Paraphrase: *The IRS was testing the issue by sending the letters, but their goal in sending the letters was to test the issue.*)

(143) The consultants maintained Eastern wouldn't generate the cash it needs *and would have to issue new debt* [Arg2-as-cond, Arg1-as-manner] **to meet its targets under the plan**. [wsj\_0475]

(Paraphrase: *If Eastern is to meet its target, it would have to issue new debt, which is the manner by which it will meet its target.*)

(144) But rather than sell new 30-year bonds, *the Treasury will issue \$10 billion of 29-year, nine-month bonds* [Result, Arg1-as-manner] **essentially increasing the size of the current benchmark 30-year bond that was sold at the previous refunding in August** [wsj\_0142]

(Paraphrase: *Issuing such bonds is the manner of increasing the size of the current benchmark 30-year bond, but it also results in increasing the size.*)

All of the senses that can co-occur with Manner can also occur on their own, so these are two separate decisions that annotators have needed to make.

**Expansion.Substitution** This tag is used when arguments are presented as exclusive alternatives, with one being ruled out.

The label EXPANSION.SUBSTITUTION.ARG1-AS-SUBST is used when Arg1 conveys the alternative which is left after the situation associated with Arg2 is ruled out.

(145) Eliminate arbitrage *and liquidity will decline* instead of **rising**, creating more volatility instead of less. [wsj\_0118]

(146) Rather than **sell 39-cents-a-pound Delicious**, *maybe we can sell 79-cents-a-pound Fujis*,” says Chuck Tryon, ... [wsj\_1128]

(147) Intervention, he added, is useful only *to smooth disorderly markets*, **not to fundamentally influence the dollar’s value**. [wsj\_0240]

The label EXPANSION.SUBSTITUTION.ARG2-AS-SUBST is used when Arg2 conveys the alternative which is left after the situation associated with Arg1 is ruled out.

(148) *He could develop the beach through a trust*, **but instead is trying have his grandson become a naturalized Mexican so his family gains direct control**. [wsj\_0300]

(149) *Nor are any of these inefficient monoliths likely to be allowed to go bankrupt*. Rather, **the brunt of the slowdown will be felt in the fast-growing private and semi-private ”township” enterprises**, ... [wsj\_1646]

(150) *However, StatesWest isn’t abandoning its pursuit of the much-larger Mesa*. StatesWest, which has a 7.25% stake in Mesa, said (Implicit=instead) **it may purchase more Mesa stock or make a tender offer directly to Mesa shareholders**. [wsj\_0321]

There is an issue as to what to include in the *Arg1* span. A situation must be explicitly signalled as being unrealized or inferred to be unrealized. Among such signals are negation, modality, *downward-entailing* predicates like *fail* or *refuse*, or *state-expecting* predicates like *plan* (Webber, 2013). We refer to this as what *licenses* ARG2-AS-SUBST. In the simplest case, the licenser is in the same clause as the unrealized situation: In Ex. 151, negation (*won’t*) is part of *Arg1*, as is the licensing modal in Ex. 152.

(151) *However, their search notably won’t include natural gas or pure methanol . . . in tests to be completed by next summer*. Instead, **the tests will focus heavily on new blends of gasoline**. [wsj\_2030]

(152) *He could develop the beach through a trust*, **but instead is trying have his grandson become a naturalized Mexican so his family gains direct control**. [wsj\_0300]

In more complex cases, the licensing element sits above the clause conveying the unrealized situation. For consistency, the licensing element is always included in *Arg1*. For example, in Ex. 153, *Arg1* includes the licensing negation over the unrealized situation of “realizing private market values by selling assets”. In Ex. 154, *Arg1* includes the licensing state-expecting predicate *expect* over the unrealized situation of “selling 44 of the most important stolen strips”. And in (155), *Arg1* includes the licensing downward-entailing predicate *denied* over the unrealized situation of “the finance ministry playing a part in the bank’s decision”.

(153) *But Santa Fe, currently trading at 18 7/8, isn’t likely to realize private market values by selling assets, . . .* **Its plan, instead, is to spin off the remainder of its real estate unit and to possibly do the same with its mining and energy assets**. [wsj\_0331]

(154) Two days earlier, his attorney met in a Park Avenue law office with a cartoon dealer *who expected to sell 44 of the most important stolen strips to Mr. Russell for \$62,800*. Instead, **New York City police seized the stolen goods**, and Mr. Krisher avoided jail. [wsj\_0450]

(155) *A Sanwa Bank spokesman denied that the finance ministry played any part in the bank’s decision*. (Implicit=instead) **“We made our own decision,”** [wsj\_1421]

### 4.3 Effect of Modifiers on Connectives: *Even*

Connectives are often found with modifiers. The set of modifiers that appear in the PDTB-3, along with the connectives they modify can be found in Appendix B.

Given the senses in the PDTB-3 hierarchy, there are two modifiers (*even* and *not*) that can change the sense of an explicit connective or AltLex phrase. Here we discuss *even*. The next section discusses *not*.

In reviewing the annotation of CONTRAST and CONCESSION tokens for the PDTB-3, we noticed that many explicit connectives modified by *even* had been assigned a second, CONCESSION sense, which did not occur with the unmodified form or with any other modifiers. For example, all tokens of *even before* were annotated ARG1-AS-DENIER as well as PRECEDENCE (the most common sense label for *before*). We therefore adopted a special convention for annotating tokens with connectives modified by *even*:

- In the case of *even if*, *even so*, *even though* and *even with*, only a CONCESSION sense is annotated;
- In the case of *even after*, *even as*, *even before*, *even if*, *even then*, *even when* and *even while*, annotators can label either just CONCESSION, or both CONCESSION and a sense associated with an unmarked version of the connective.

### 4.4 Effect of Explicit and Implicit Negation

**Connectives modified by Explicit Negation** Two of the senses in the PDTB-3 hierarchy (Table 1) are opposite to each other — namely, SIMILARITY and CONTRAST. When a connective which normally expresses one of these senses (e.g., *like*) is modified by *not*, it flips the sense to its opposite: The following token from wsj\_0296 is labelled SIMILARITY, while the one from wsj\_2232 is labelled CONTRAST.

(156) *Her Susie Diamond handles a song the way the greats do, like she’s hearing the way it should sound inside her head and she’s concentrating on matching that internal tone.* [wsj\_1831]

(157) *There wasn’t a lot of panic selling, either domestically or internationally. Not like Friday where they just took {the market} apart.* [wsj\_2232]

With other connectives, *not* does not change the sense.

**Connectives embodying Implicit Negation: *Without*** When the connective *without* is used to express MANNER, its implicit negation needs to be interpreted as the manner that is not employed, even though there is no explicit lowering of negation into Arg2. For example, (158) can be paraphrased as “with not scrimping on its existing brands”.

(158) Without P&G’s backing, Noxell might not have been able to spend the estimated \$5 million to \$7 million needed to accomplish that without scrimping on its existing brands. [wsj\_0589]

**AltLex expressions embodying Implicit Negation** Implicit negation in an AltLex expression may or may not be associated with a negative sense. In the case of *barring*, relations have been labelled with the negative sense ARG2-AS-NEGCOND, as in

(159) *it expects a “substantial increase” in consolidated profit for the full year, barring major currency rate changes* [wsj\_2089]

. In other cases, implicit negation may still be associated with a positive sense — for example, *preventing* whose use is still associated with RESULT, as in

(160) *the crystal changes his team introduced apparently pins the magnetic fields in place, preventing them from lowering current-carrying capacity [wsj\_0123]*

## 5 Intra-sentential Discourse Annotation

While the PDTB-2 contains 40600 discourse relations, some syntactic and textual contexts were not annotated, due to the guidelines for identifying discourse relations. In particular, PDTB-2 guidelines limited annotation to (a) Explicit relations lexicalized by discourse connectives, and (b) Implicit, AltLex, EntRel and NoRel relations between paragraph-internal adjacent sentences and between clauses within sentences separated by colons or semi-colons. Further, explicit discourse connectives were drawn from the pre-defined syntactic classes of subordinating conjunctions, coordinating conjunctions, and discourse adverbials. And strict constraints were placed on how a relation’s arguments could be realized syntactically: with a few exceptions, arguments had to be realized as one or more *clauses* or *sentences*.

Defining the scope of the annotation in this way, however, precluded consideration of a wider set of discourse relations. This has been addressed in the PDTB-3, where  $\sim$ 13k additional intra-sentential relations have been annotated, including free adjuncts (discussed in Section 5.1), free to-infinitives (Section 5.2), prepositional clausal subordinates (Section 5.3), conjoined verb phrases (Section 5.4), implicitly conjoined clauses (Section 5.5), and marked syntactic constructions (Section 2.7.2).

For all but marked syntactic constructions (AltLexC), potentially relevant tokens were automatically identified (Prasad et al., 2015) using a combination of Penn Treebank (Marcus et al., 1993) and Propbank (Palmer et al., 2005) annotations. Subordinated clause contexts were identified by searching for adjunct (ARGM) arguments of verbs in Propbank, and then restricting the result to clausal arguments using an alignment of Propbank ARGM arguments with Penn Treebank (PTB) syntactic structures. Further heuristics applied to these structures enabled the tokens to be divided into separate sets, creating distinct subtasks corresponding to free adjuncts, TO-infinitives and preposition-subordinated clauses. To facilitate future analysis, tokens retain the ARGM semantic role label they were found with, as well as the target verb that the ARGM clause modifies. For example, the free adjunct annotation for Ex. 163 (see Section 5.1) also indicates that the **Arg2** free adjunct modifies the verb *open* in Propbank and carries the Propbank semantic role ARGM-ADV.

VP conjunction contexts in PDTB-3 were identified with heuristics applied solely to the PTB, and were also annotated as a separate subtask.

Further filtering was done for each of the intra-sentential subsets, to remove tokens that (a) were not appropriate to annotate, or (b) were already annotated in PDTB-2. The final set of tokens provided to annotators were pre-annotated with the Arg2 of the relation as extracted from the PTB and Propbank, as applicable, but the annotators were encouraged to adjust the Arg2 span, if needed, as they completed the annotation of the relation.

One consequence of annotating implicit relations in VP conjunction contexts was a decision to revisit *Intra-S* S-conjunction contexts annotated in PDTB-2, since they allow similar inferences of additional implicit relations. This is illustrated in Examples (161-162), where an implicit discourse relation (CONCESSION.ARG2-AS-DENIER) is inferred in addition to the explicit CONJUNCTION. Section 5.5 describes the augmentation of PDTB-2 S-conjunction contexts with these additional implicit relations.

- (161) John Rowe, president and chief executive officer of New England Electric, said the company’s return on equity could suffer if *it made a higher bid* and **its forecasts related to PS of New Hampshire – such as growth in electricity demand and improved operating efficiencies – didn’t come true** (EXPANSION.CONJUNCTION) [wsj\_0013]
- (162) John Rowe, president and chief executive officer of New England Electric, said the company’s return on equity could suffer if *it made a higher bid* and Implicit=yet **its forecasts related to PS of New Hampshire – such as growth in electricity demand and improved operating efficiencies – didn’t come true** (COMPARISON.CONCESSION.ARG2-AS-DENIER) [wsj\_0013]

In the rest of this section, we discuss distinctive characteristics and annotation of each of the intra-sentential contexts.

## 5.1 Free Adjuncts (FAs)

Free adjuncts (FAs) are syntactically subordinated clauses without any explicit subordinating conjunction, such as the FA in Ex. 163.

- (163) *Treasurys opened lower*, Implicit=as a result of **reacting negatively to news that the producer price index – a measure of inflation on the wholesale level – accelerated in September.** (CONTINGENCY.CAUSE.REASON)[wsj\_2428]

For automatic identification and pre-annotation of potential FA contexts, a span had to be parsed in the PTB as a free subordinate clause and labeled in Propbank as either ARGM-ADV (adverbial), ARGM-CAU (causal), ARGM-MNR (manner), ARGM-PNC/PRP (purpose), ARGM-PRD (secondary predication), or ARGM-TMP (temporal). Following the PDTB-3 argument naming convention (Section 3.1), an FA is Arg2 because it is syntactically subordinated, and its matrix clause is Arg1.

A few of the spans found this way contained explicit connectives, and so had already been annotated in the PDTB-2 (Ex. 164). Other spans contained verb-derivative forms such as “given” (Ex. 165) and “depending on” (Ex. 166) which functioned essentially as discourse connectives and so were labelled as explicit connectives.

- (164) *Sony, for its part, could decide that the cost of a Warner settlement or court fight is too high*, **choosing instead to find someone else to run Columbia, ...** (EXPANSION.SUBSTITUTION) [wsj\_0578]
- (165) Some dealers said *the Treasury’s intent is to help government bond dealers gauge investor demand for the securities*, given **uncertainties about when the auction will occur.** (CONTINGENCY.CAUSE.REASON) [wsj\_0142]
- (166) *The price of the second simulator ranges between C\$16.4 million and C\$18 million*, CAE said, depending on when the Army exercises its option. (CONTINGENCY.CONDITION.ARG2-AS-CONDITION) [wsj\_2118]

Spans that did not contain an explicit connective and that expressed a discourse relation with their matrix clause were annotated as either Implicit (Ex. 163) or AltLex (Ex. 167) relations. Spans that expressed a relation with an entity in the matrix clause were labelled EntRel (Ex. 168).

- (167) *The company is operating under Chapter 11 of the federal Bankruptcy Code*, giving it court protection from creditors’ lawsuits while it attempts to work out a plan to pay its debts. (CONTINGENCY.CAUSE.RESULT) [wsj\_0031]

(168) “*I am happy to see the spirit of the people,*” said Mr. Sisulu, (EntRel) **looking dapper in a new gray suit.** (wsj\_2454)

### 5.1.1 Implicit connectives in FAs

PDTB guidelines suggest that “implicit connectives” be semantically appropriate and sound syntactically natural in context. Most FA contexts accept typical implicit connectives, including bare adverbials (e.g., *thus, instead*) and bare subordinating conjunctions (e.g., *while, after*), In other contexts, however, only a simple prepositional subordinator such as *in, by, for, with*, etc. (Ex. 169-170) or a complex prepositional subordinator, such as *as a result of, instead of, because of* (Ex. 171) sound natural.

(169) *In October doctors were debating the product’s safety,* (Implicit=with) **some claiming it caused infections.** (EXPANSION.INSTANTIATION.ARG2-AS-INSTANCE) [wsj\_2144]

(170) The first hybrid corn seeds *produced* (Implicit=by) **using this mechanical approach** were introduced in the 1930s and they yielded as much as 20% more corn than naturally pollinated plants. (EXPANSION.MANNER.ARG2-AS-MANNER) [wsj\_0209]

(171) *Treasurys opened lower,* (Implicit=as a result of) **reacting negatively to news that the producer price index – a measure of inflation on the wholesale level – accelerated in September.** (CONTINGENCY.CAUSE.REASON) [wsj\_2428]

Because FAs often ellipse relevant auxiliary verbs, auxiliaries have often been added to the implicit connective for naturalness, as in Ex 172, even though the auxiliary is not technically part of the connective.

(172) (As a result of being) **Stung by the Giuliani ads,** *Mr. Dinkins’s TV consultants, Robert Shrum and David Doak, finally unleashed a negative ad of their own.* (CONTINGENCY.CAUSE.REASON) [wsj\_0041]

Note that the implicit connective should be natural with respect to any alterations in the Arg2 span, which may have been changed from its pre-annotated free adjunct. For example, the attribution verb *alleging* has been excluded from Arg2 in Ex. 173, so the implicit connective *because of* sounds natural with respect to what has been annotated as Arg2.

(173) *In addition, Lionel began a lawsuit in federal District Court in New York seeking to enjoin the offer,* alleging, (Implicit=because of) **among other things, violations of federal securities law and fraudulent manipulation of the market for Lionel’s securities.** (CONTINGENCY.CAUSE.REASON) [wsj\_1917]

Implicit connectives annotated on FAs vary with fine-grained distinctions in their syntax.

**FAs with verb forms** The majority of FAs involve a verb complex and an implicit subject that is generally the same as the matrix clause, though it may be the speaker/writer.

When the verb complex includes a present participle, natural sounding implicit connectives include adverbials (Ex. 174) and subordinating conjunctions (Ex. 175), as well as simple and complex prepositional subordinators (Ex. 170-171).

(174) *He now runs Policy Communications in Washington,* (Implicit=specifically) **consulting to media companies.** (EXPANSION.LEVEL-OF-DETAIL.ARG2-AS-DETAIL) [wsj\_2451]

- (175) For Ms. Cunningham the architectural discombobulation matched the discrepancy *she felt* (Implicit=when) **living in the AnaMor Towers as a little girl.** (TEMPORAL.SYNCHRONOUS) [wsj\_2343]

When the verb complex contains a past participle, natural sounding connectives include bare connectives (Ex. 176–177) and connectives followed by auxiliary verbs inflected for the appropriate tense and aspect endings (Ex. 178).

- (176) (Implicit=When) **Turned loose in Shane Longman’s trading room,** *the yuppie dealers do little right.* (TEMPORAL.SYNCHRONOUS) [wsj\_2402]

- (177) (Implicit=If) Left to its own devices, index arbitrage will become more and more efficient, making it harder and harder to do profitably. (CONTINGENCY.CONDITION.ARG2-AS-COND) [wsj\_0118]

- (178) (As a result of being) **Stung by the Giuliani ads,** *Mr. Dinkins’s TV consultants, Robert Shrum and David Doak, finally unleashed a negative ad of their own.* (CONTINGENCY.CAUSE.REASON) [wsj\_0041]

**FAs with no verb form** The Arg2 associated with a FA can have a non-verbal predicate (noun, adjective, or prepositional phrase) if the FA lacks a verb complex (Ex. 180) or if its verb is excluded as the attribution of Arg2 (as *citing* is in Ex. 179 or as *alleging* is in Ex. 173). Typical implicit connectives here include connectives/subordinators followed by the preposition *of* and optional auxiliaries conveying tense and aspect:

- (179) *The Nicaraguan president,* citing (Implicit=as a result of) **attacks by the U.S.-backed rebels,** *suspended a 19-month-old cease-fire ...* (CONTINGENCY.CAUSE.REASON) [wsj\_0174]

- (180) (Implicit=As a result of being) **Unable to unload UAL and other airline shares,** *takeover-stock speculators, or risk arbitrageurs, dumped every blue-chip stock they had.* (CONTINGENCY.CAUSE.REASON) [wsj\_0118]

**FAs with explicit subjects** Some FAs have an explicit subject, with or with a verb complex. Typical implicit connectives here include the subordinator *with* or bare connectives followed by *of*.

- (181) (Implicit=As a result of) **The world being the nasty place it is,** *we want Presidents to have the freedom to order operations in which someone might get killed.* (CONTINGENCY.CAUSE.REASON) [wsj\_1307]

- (182) (Implicit=With) **Inflation surging,** *the pound began falling against the mark.* (CONTINGENCY.CAUSE.REASON) [wsj\_0571]

### 5.1.2 AltLex phrases in FAs

In several FAs, the relevant discourse relations are expressed through a combination of the FA verb complex and the anaphoric reference to Arg1 with the FA implicit subject. In these cases, an AltLex phrase has been identified.

- (183) *In countries such as Taiwan, South Korea and Singapore, economies are growing,* **resulting in a rise in disposable income that consumers can use for soft drinks.** (CONTINGENCY.CAUSE.RESULT)[wsj\_0245]

In some FAs, the relevant discourse relation is expressed as a syntactic construction. In such cases, the entire Arg2 has marked as the alternative lexicalization (cf. Section 2.7.2).

- (184) **Crude as they were**, *these early PCs triggered explosive product development in desktop models for the home and office.* (COMPARISON.CONCESSION.ARG1-AS-DENIER) [wsj\_0022]
- (185) **Had he been a little less gung-ho**, *”I’d have gotten the thing on the ground and headed for the nearest bar,”* Mr. Brown says. (CONTINGENCY.CONDITION.ARG2-AS-COND)[wsj\_1394]

Note that many of these AltLex phrases contain the same kind of modifiers that appear with Explicit connectives (cf. Appendix B), and even with Implicit connectives (cf. Section 5.2), such as *partly*, *possibly*, *probably*, *largely*, and *in part*:

- (186) *Each day that Congress fails to act . . . will cause additional disruption in our borrowing schedule* **possibly resulting in higher interest costs to the taxpayer.** [wsj\_0075]
- (187) *Beazer PLC, a major British building materials and construction concern, reported a 24% jump in pretax profit for its latest financial year,* **helped largely by contributions from its U.S. unit, Koppers Co.** [wsj\_1921]
- (188) *Federal’s stock price, however, has held up well,* **driven in part by the general run-up of airline stocks.** [wsj\_1394]

While no special annotation was done on these modifiers, they can be found in Field 1 (Conn Spanlist) of relevant AltLex tokens, in case a researcher wants to carry out further analysis (cf. Section 8.1).

## 5.2 Infinitival Clauses

Infinitival clauses are syntactically subordinated clauses headed by a *to*-infinitive. Their automatic identification and pre-annotation required that a span be parsed in the PTB as a free *to*-infinitive clause and labelled in PropBank as an ARGM-PRP or ARGM-PNC adjunct. *Free* means that there is no subordinating element introducing the clause. Annotators were told to annotate *in order* as an implicit connective, so free *to*-infinitives have been annotated as implicit relations, while explicit *in order to*-infinitives have been annotated as explicit relations. *To*-infinitives with subjects like “for John to eat the cake”, have been extended to include the subject. (Those Propbank ARGM-PNC roles whose clausal arguments were headed by *so*, *so that* and *because of* rather than *to*-infinitives had either been annotated in the PDTB-2 or were annotated as part of a separate task.)

As with other pre-annotated spans, in order to be taken to be **Arg2** of a discourse relation, its **Arg1** had to be a separate distinct action. If not, the pre-annotated span was rejected, as in

- (189) The move is designed to ward off a hostile takeover attempt by two European shipping concerns, Stena Holding AG and Tiphook PLC.

Here, the pre-annotated span *to ward off a hostile takeover attempt by two European shipping concerns . . .* was rejected as part of a discourse relation, because the span *The move is designed* is not a separate distinct action.

Pre-annotated spans were also rejected if the *to*-infinitive served as a complement or reduced relative clause (RRC) on a noun phrase (NP), as in

- (190) . . . dozens of additional crews and transport aircraft were on alert awaiting orders to move emergency supplies

where *to move emergency supplies* is a complement on *orders*.

In total, over 2000 clausal adjuncts playing ARGM-PRP or ARGM-PNC roles in PropBank were eventually included in the PDTB-3 as **Arg2** of a discourse relation — the majority as implicit relations (over 1700), but some as explicit relations, AltLex or AltLexC relations, or entity relations (EntRel). While the sense ascribed to these relations was most often PURPOSE.ARG2-AS-GOAL, many were taken to have a different sense, such as CAUSE.RESULT, CAUSE.REASON or CONDITION.ARG2-AS-COND. Many of the relations have a second sense annotated as well — usually MANNER or SUBSTITUTION. These are described in the next sections, contrasting tokens annotated as PURPOSE with ones annotated as RESULT, tokens annotated as PURPOSE with ones annotated as CONDITION, and tokens annotated with multiple senses, as opposed to just a single sense.

**purpose vs result** PURPOSE indicates why an action was undertaken; RESULT indicates its consequence. Generally PURPOSE only makes sense for actions that have a volitional agent, as in

(191) The Galileo project started in 1977, and *a number of project veterans were on hand* **to watch the launch**. [wsj\_1817]

Such tokens can usually have the phrase *in order* inserted as their implicit connective. This is not the case for tokens sense-labelled as RESULT, such as (192), where *therefore* is more appropriate as an implicit connective.

(192) *Georgia Gulf stock rose \$1.75 a share yesterday* **to close at \$51.25 a share** [wsj\_0080]

**purpose vs condition** The conditional sense ARG2-AS-COND has been annotated when the situation specified in *Arg1* holds if the situation specified in **Arg2 might be** the agent’s purpose, but there is no assertion that it **is**. As such, while *in order* is a possible implicit connective, a better alternative is an *if*-clause prefix such as *if it is, if they are, if one were*, etc.), as in

(193) *Banks need a competitive edge* **to sell their products**. [wsj\_0238]  
Paraphrase: *Banks need a competitive edge if they are to sell their products*

(194) He said *the index would have to be in the low 40% range for several months* **to be considered a forecast of recession**. [wsj\_0036]  
Paraphrase: *the index would have to be in the low 40% range for several months if it is to be considered a forecast of recession*

.

One signal for a conditional sense is the presence of a modal (e.g., “need”, “have to”, “must”, or “require”) or future tense (or present tense used as future) in *Arg1*. However, this doesn’t hold if the modal or future operator scopes the entire relation, meaning it can be extracted, leaving both *Arg1* and *Arg2* in its scope, as in

(195) The two companies have been discussing a transaction *under which Fresenius would buy Delmed stock for cash* **to bring its beneficial ownership to between 70% and 80% of Delmed’s fully diluted common stock**. [wsj\_1066]

Since *would* can be extracted to scope both arguments (i.e., *that would have Fresenius buys Delmed stock for cash to bring its beneficial ownership to . . .*), a purpose sense is appropriate here (ARG2-AS-GOAL). In general, if a modal or future in *Arg1* can be taken to scope both arguments, then the sense relation has been taken to be either PURPOSE or RESULT, depending on criteria noted previously.

Note that the conditional sense ARG2-AS-COND may also be appropriate if *Arg1* contains negation or is a question. To decide, one should consider a positive paraphrase of *Arg1* in the case of negation, while considering a non-question paraphrase in a question context, as in

(196) . . . which, unlike utilities, aren't regulated *and therefore don't need government approval to construct new plants*. [wsj\_0560]

(197) *Do you really need this much money to put up these investments?* [wsj\_0629]

An appropriate positive paraphrase of Example 196 is “they need government approval if they are to construct new plants”, hence ARG2-AS-COND is an appropriate label. For Example 197, “you really need this much money if you are to put up these investments” was not considered an appropriate non-question paraphrase, while “you really need this much money in order to put up these investments” was. Hence ARG2-AS-GOAL is the appropriate sense label.

In a few cases, annotators have taken the infinitival to provide DETAIL about its matrix clause (ARG2-AS-DETAIL), either as the only relation that holds (Example 198), or along with PURPOSE (Example 199). So this is another relation to consider.

(198) . . . *expanding the monthly reports on program trading to cover specific days or even hours of heavy program trading and who was doing it*. [wsj\_0349]

(199) Rep. Hamilton said *the bill will be modified substantially to call for two meetings each year . . .* [wsj\_2082]

**Multiple Senses** As already noted, annotators often infer more than one relation between an infinitival and its host — most often, ARG1-AS-MANNER, but sometimes ARG2-AS-SUBST or ARG2-AS-DETAIL.

To assess whether ARG1-AS-MANNER also holds, a good test is whether *thereby* can be inserted as a second implicit connective. Example 200 shows PURPOSE tokens, where an additional ARG1-AS-MANNER sense has been inferred in the case of (a), but not (b). Similarly, example 201 shows CONDITION tokens with and without an additional ARG1-AS-MANNER sense, and example 202 shows a similar pair for RESULT.

(200) a. **To transfer information from one to the other**, *employees make printouts and enter the data manually*. [wsj\_0445]

b. *Due to the earthquake in San Francisco, Nissan is donating its commercial air time to broadcast American Red Cross Emergency Relief messages*. [wsj\_0453]

(201) a. But Mr. Tonkin said *dealers should slash stocks to between 15 and 30 days to reduce the costs of financing inventory*. [wsj\_0618]

b. Further, he said, *the company doesn't have the capital needed to build the business over the next year or two*. [wsj\_0092]

- (202) a. *On most days, the desert’s heat and the cool of the ocean combine to create a mist like a damp rag.* [wsj\_1121]  
 b. *West Germany’s 7% issue due October 1999 rose 0.13 point to 99.93 to yield 7.01%.* [wsj\_1213]

To assess whether ARG2-AS-SUBST also holds, a good test is whether *instead* can be inserted as a second implicit connective. Instances in the PDTB-3 are domain-specific, all involving predication in *Arg1* of an agent *leaving, resigning* or *giving up*, for the PURPOSE of doing something else – e.g.,

- (203) *Mr. Corr resigned to pursue other business interests*, the airline said. [wsj\_0509]  
 (204) *Jeanette Traverso, a California lawyer, gave up running three times a week to play a weekly round of golf . . .* [wsj\_0409]

**Infinitivals and modifiers** Infinitivals co-occur with words/phrases that are considered modifiers when the co-occur with explicit connectives (e.g., *partly, mainly, apparently, only* — cf. Appendix B), as in

- (205) A broad rally began when several major processors began *buying futures contracts, apparently to take advantage of the price dip.* [wsj\_2265]  
 (206) **Partly to help clear the myriad obstacles facing any overseas company trying to penetrate Japan**, *tiny Candela turned to Mitsui & Co., one of Japan’s largest trading companies, for investment.* [wsj\_0083]  
 (207) *Thrifts continued to shed assets in August, mainly to comply with stiffer capital rules under the S&L bailout law.* [wsj\_1364]

However, since bare *to*-infinitives only participate in implicit relations, these modifiers have not be acknowledged except in one case noted earlier (Section 2.6), where the sequence *only to* needs to be treated as distinct single connective.

### 5.3 Other Subordinated Structures

The search for subordinated clause contexts over the PTB and Propbank also yielded structures other than free adjuncts and *TO*-infinitives. These include, primarily, subordinate clauses introduced by prepositional subordinators (Ex. 208–211), which were then annotated as explicit connectives.

- (208) With membership of the Church of England steadily dwindling, *strong-willed vicars are pressing equally strong-willed and often non-religious ringers to attend services.* (CONTINGENCY.CAUSE.REASON) [wsj\_0089]  
 (209) 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. (COMPARISON.CONCESSION.ARG2-AS-DENIER) [wsj\_0013]  
 (210) He also said that the group reduced its offer *because it wasn’t allowed to see Georgia Gulf’s confidential financial information without agreeing that it wouldn’t make an offer unless it had Georgia Gulf’s consent.* (CONTINGENCY.NEGATIVE-CONDITION.ARG2-AS-NEGCOND) [wsj\_0080]

(211) Eliminate arbitrage and *liquidity will decline* instead of **rising**, creating more volatility instead of less. (EXPANSION.SUBSTITUTION.ARG1-AS-SUBST) [wsj\_0118]

Another major class of constructions led to their annotation as AltLexC, a subtype of AltLex where the relation is expressed by the syntactic construction, for example with the predicate and AUX inversion in Exs. (212-213).

(212) **Crucial as these elections are for Greece**, *pressing issues of state are getting lost in the shuffle*. (COMPARISON.CONCESSION.ARG1-AS-DENIER) [wsj\_2431]

(213) *On the other hand*, **had it existed then**, *Cray Computer would have incurred a \$20.5 million loss*. (CONTINGENCY.CONDITION.ARG2-AS-COND) [wsj\_0018]

## 5.4 Conjoined Verb Phrases

The two-step process for pre-annotating conjoined VPs prior to annotation involved:

- Searching the PTB parses for all sister VPs separated by a conjunction, conjunction phrase (e.g. *rather than*) or punctuation, and an optional adverbial.
- For each such pair of sisters, pre-annotating the righthand VP as Arg2 of a potential discourse relation. The relation type will be **Explicit** if a conjunction or conjunction phrase appears between the sisters. Sisters separated by punctuation have relation type **Implicit**, which may change to **AltLex** during sense annotation if material in either *Arg1* or **Arg2** makes the insertion of an implicit connective seem redundant.

The set of tokens caught this way was further filtered to eliminate cases of *Argument/adjunct cluster coordination* (Mouret, 2006; Steedman, 1989, 2000), also called *non-constituent conjunction* in which there is no verb in the righthand conjunct, even though labelled a VP in the Penn TreeBank. So in (214), corresponding spans consisting of a direct and indirect objects of *pay* have been coordinated as sister VPs, as have corresponding spans consisting of an adverb and a PP in (215).

(214) “I pay a lot to the farmer and **five times the state salary to my employees**,” he says [wsj\_1146]

(215) She adopted 12 of assorted races, naming them the Rainbow Tribe, and driving her husband first to despair and **then to Argentina**. [wsj\_1327]

Such constructions involving verbless VPs are relatively easy to recognize and exclude from the set.

Some conjoined VPs have been parsed in the Penn TreeBank as conjoined S-nodes with null subjects. While they were missed in the above process, they were later caught when intra-sentential conjoined clauses were identified for potential annotation (Section 5.5).

This pre-annotation process identified conjoined tensed VPs (Ex. 216), conjoined adjunct VPs (Ex. 217), and conjoined imperative sentences (Exs. 218–219), all of which have been parsed in the Penn TreeBank as conjoined VPs.

(216) It employs 2,700 people and **has annual revenue of about \$370 million**. [wsj\_0007]

(217) But many owners plan to practice frugality – crossing out the old code and **writing in the new one** until their stock runs out. [wsj\_1270]

(218) Then take the expected return and **subtract one standard deviation**. [wsj\_1564]

(219) Be careful boys; **use good judgment**. [wsj\_0596]

Because conjoined VPs might also have discourse adverbials conveying an additional relation between the conjuncts — adverbials which will have already been annotated in the PDTB-2, such tokens were imported from the PDTB-2, with sense labels automatically updated to reflect the revised relation hierarchy, to help in the new annotation of the conjoined VPs.

As noted in Section 3.1, for coordinating conjunctions, annotators were asked to annotate just the conjuncts — here, just conjoined VPs. Thus, in Ex. 39, *Arg1* should subsume only *joined the firm in 1963*, and not the subject *he*. The same goes for Ex. 216.

Annotators were also asked to omit material that belongs semantically to both conjuncts of a conjoined VP. This too may mean that the spans of VP conjuncts in the PDTB-3 may not completely match the projection of structures in the Penn TreeBank. For example, in

(220) UAL ...reversed course and plummeted in off-exchange trading after the 5:00 p.m. EDT announcement. [wsj\_1305]

the PTB parses *reversed course* as conjoined with *plummeted in off-exchange trading after the 5:00 p.m. EDT announcement*, even though both events happen *in off-exchange trading after the 5:00 p.m. EDT announcement*. In the PDTB-3, **Arg2** is taken simply to be *plummeted*.

With conjoined VPs, the spans of the arguments should be of the same type — both tensed clauses, free adjuncts, bare *to*-infinitives, etc. So in Ex. 221, since **Arg2** is the *to*-infinitive *then to draft medical-practice guidelines*, *Arg1* should be the *to*-infinitive *to measure the effectiveness of medical treatments*.

(221) The NAM embraces efforts, ..., *to measure the effectiveness of medical treatments* and **then to draft medical-practice guidelines**. [wsj\_0314]

Finally, when annotating conjoined VPs, the only attributions that should be retained are ones that contribute to the semantics of the relation (as in Ex. 223, where the PURPOSE of declaring something a pesticide is so that it can be pulled from the marketplace). In contrast, in Ex. 222, neither *said* nor *added* contribute to the annotation of CONCESSION, so should be omitted from the corresponding spans of *Arg1* and **Arg2**.

(222) The company, based in San Francisco, said *it had to shut down a crude-oil pipeline in the Bay area to check for leaks* but added that **its refinery in nearby Richmond, Calif., was undamaged**. [wsj\_1884]

(223) Give the EPA more flexibility to *declare a pesticide an imminent hazard* and **pull it from the marketplace**. [wsj\_0964]

If the same relation is interpreted as holding either between attribution phrases or between their complements, the relation should be annotated between the latter, indicating the minimal spans that give rise to the particular relational sense.

**Conjoined VPs with a single sense relation** Like any explicit connective, an explicit conjunction can have a one or more senses. In the case of Ex. 224–226, the connective conveys a single sense.

- (224) Chemical named James Kennedy, . . . , *to assume Mr. Edelson’s duties* and **to be trading manager for derivative products, including swaps and interest-rate options** (EXPANSION.CONJUNCTION) [wsj\_0304]
- (225) These active suspension systems *electronically sense road conditions* and **adjust a car’s ride** (CONTINGENCY.PURPOSE.ARG2-AS-GOAL) [wsj\_0956]
- (226) Stocks *closed higher in Hong Kong, Manila, Singapore, Sydney and Wellington*, but **were lower in Seoul**. (COMPARISON.CONTRAST) [wsj\_0231, wsj\_0374]

**Conjoined VPs with multiple sense relations** As for multiple sense relations between conjoined VPs, these can arise in two ways. First, there can be a distinct explicit connective in the second conjunct that expresses a separate relation between the conjoined VPs. (Many of these connectives will have been already been annotated in PDTB-2 and imported to help in the annotation of the conjoined VPs.)

- (227) a. A Chemical spokeswoman said *the second-quarter charge was “not material”* and (EXPANSION.CONJUNCTION) **that no personnel changes were made as a result** [wsj\_0304].
- b. A Chemical spokeswoman said *the second-quarter charge was “not material”* **and that no personnel changes were made** as a result (CONTINGENCY.CAUSE.RESULT) [wsj\_0304].

Second, another relation may be inferred as holding between the conjuncts, in addition to that signalled by the explicit conjunction, as in Ex. 228-229.

- (228) We’ve got to *get out of the Detroit mentality* **and** Implicit=instead **be part of the world mentality**, declares Charles M. Jordan, GM’s vice president for design . . . [wsj\_0956]  
(EXPANSION.CONJUNCTION, EXPANSION.SUBSTITUTION.ARG2-AS-SUBST)
- (229) . . . Exxon Corp. *built the plant* **but** (Implicit=then) **closed it in 1985**. [wsj\_1748]  
(COMPARISON.CONCESSION.ARG2-AS-DENIER, TEMPORAL.ASYNCHRONOUS.PRECEDENCE)

As noted in Section 8.5, the new convention in the PDTB-3 is to create a separate annotation token for the implicit relation between the two conjuncts, with an appropriate implicit connective and appropriate sense labels. Argument spans of the explicit and the implicit relation are not required to be the same, so the spans of the new token can be adjusted if needed.

**AltLex in Multi-sense Conjoined VPs** If another relation is inferred as holding in a conjoined VP, in addition to that signalled by the explicit conjunction, but inserting an implicit connective is perceived as redundant, appropriate material can be annotated as *AltLex*, as in Ex. 230.

- (230) His policies *went beyond his control* **and resulted . . . in riots and disturbances**. [wsj\_0290]  
(EXPANSION.CONJUNCTION, CONTINGENCY.CAUSE.RESULT)

## 5.5 Implicit Conjoined Clauses

While intra-sentential clauses explicitly conjoined with *and* and *but* were annotated in the PDTB-2, fewer implicitly conjoined tokens were so annotated – only those in which the clauses were connected by punctuation such as semi-colon, colon, or dashes. Suspecting that these were not all the implicitly conjoined clauses in the PTB, other S-siblings were examined — in particular, ones separated by a non-lexical element such as comma, semi-colon, colon, dash, double-dash, question mark, period or no marking. This revealed 1205 tokens, some of which matched or overlapped with those already annotated in PDTB-2. Specifically, 369 of those tokens were an exact match (both arguments) with one of the 1205 tokens; 163 were a partial match in terms of argument spans, while 62 PDTB-2 tokens were NOT matched, due to having one or both arguments not labelled as an S-node in the syntax. Counting partial matches as matches, 532 of the 1205 tokens found by searching the PTB had already been annotated in PDTB-2.

A subsequent, less constrained search over parse trees in the PTB yielded over 5510 tokens. Of these, two contained S-nodes which (for some reason) had a span of zero, leaving 5508 tokens. Of these, 2964 were found to contain explicit connectives, and so were already annotated in the PDTB-2. This left 2544 tokens, including 1093 that matched tokens from the earlier set of 1205 tokens. Of the 112 unmatched tokens, five had already been annotated as explicit connectives, leaving 107 tokens. These 107 tokens were added to the 'clean' set of 2544 tokens to give 2651 tokens. However, 271 of these had already been extracted as free adjuncts (cf. Section 5.1) and could be removed, while others turned out to pair an attribution phrase with what was attributed to the speaker. The whole process left 2235 tokens to potentially be annotated as implicit conjoined clauses — some with automatically generated pre-annotation of one or both arguments, some with arguments already annotated in the PDTB-2. In both cases, annotators were asked to consider whether to accept or modify the spans, and to accept or modify the senses (if annotated in the PDTB-2).

### 5.5.1 Argument Span Annotation

When both argument spans had been automatically annotated (based on parse trees in the Penn TreeBank or imported from the PDTB-2), annotators could either accept them as is, or modify them. When only Arg2 was pre-annotated, it was because the relation is implicit. Annotators first needed to check whether there is some independent clause to serve as Arg1. If not, the token should be rejected, as in

(231) The September index fueled speculation, damaging to the dollar, **that the Federal Reserve soon will ease monetary policy further.** [wsj\_1731]

where the pre-annotated **Arg2** is the NP-complement of *speculation*. If there is an independent *Arg1*, then annotation can proceed as for any implicit relation.

There are, however, some aspects of annotation specific to sentence-internal conjoined clauses. First, anything that applies to both conjuncts should be excluded from both *Arg1* and **Arg2**, as in

(232) By October, however, *market conditions had deteriorated* **and the reset notes were targeted to be offered at a yield of between 13 1/4% and 13 1/2%.** [wsj\_0351]

The same token also shows that any discourse connective prior to *Arg1* should also be excluded from the argument spans. Secondly, unlike conjoined VPs (Section 5.4), verb-less argument spans are permitted, as long as the spans can be understood as 'clauses with elided verbs' (as in the gapped construction in

Ex. 233 or the *small clauses* in Ex. 234), 'nominalizations', or 'anaphoric expressions denoting abstract objects', or particles appearing as responses to questions.

(233) The reports, attributed to the Colombian minister of economic development, said *Brazil would give up 500,000 bags of its quota and Colombia 200,000 bags*, the analyst said. [wsj\_0437]

(234) a. *Delivery of the first aircraft is set for early November, (Implicit=and) a second for December* and two for April 1990.

b. Delivery of the first aircraft is set for early November, *a second for December and two for April 1990*. [wsj\_0243]

A third point specific to annotating implicitly conjoined clauses involves “:”, where annotators should reject a token such as that in Ex. 235, where what follows the “:” is not independent of what comes before the “:”

(235) But nobody did such a piece, reflecting a contemporary axiom: the more a scandal has to do with a congressman’s duties as a congressman, the less likely it is to catch the fancy of a network.

In contrast, what is after the colon in 236 can stand as independent of what comes before, and hence has been annotated.

(236) But *the sponsors have an explanation: “**Through the character associated with a snail**” they say, “**important values such as harmony with nature and aspirations for the future are sought.**”* [wsj\_0956]

With respect to response particles, where the response to a question comprises only a response particle like *yes* or *no*, the particle itself can be annotated as an argument, with the preceding question optionally annotated as supplementary material, as in

(237) Are such expenditures worthwhile, then? *Yes if targeted.*[wsj\_2412]

(238) Is he a victim of Gramm–Rudman cuts? *No but **hes endangered all the same*** [wsj\_0528]

Finally, we have not annotated sequences of fragments, such as the comma-conjoined fragments in

(239) 3COM CORP., Santa Clara, maker of computer communications systems, annual sales of \$386 million, OTC, slight structural damage to headquarters, communications systems already fully operational. [wsj\_1915]

### 5.5.2 Sense Annotation

Implicitly conjoined clauses have a similar range of senses as explicitly conjoined clauses. The only difference comes with non-restrictive relative clauses. Sentential relatives as in Ex. 240–241 should always be annotated.

(240) *Prices fell through levels regarded as important support areas, **which added to the selling.*** [wsj\_0437]

(241) *The best team won this Series, **which is more unusual than it may sound*** [wsj\_0561]

Non-restrictive relative clauses on NPs are annotated either as implicit relations, as in Ex. 242, or as EntRel (Ex. 243).

(242) *Takeover speculators – who have already taken a record loss estimated at more than \$700 million on UAL – started selling other stocks as well as S&P futures in an attempt to hedge against a further UAL blood bath.* [wsj\_1208]

(243) *Free markets, free minds and free elections have an appeal that seems to get muddled only when delivered through U.N. organizations – which of course are made up largely of governments that fear these principles at home.* [wsj\_0208]

One final convention: When Arg1 introduces an entity (as in Ex. 244), the relation to an Arg2 that describes that entity should be labelled as an implicit ARG2-AS-DETAIL (not ENTREL), because Arg1 doesn't do anything more than introduce the entity.

(244) *First came his predictable fusillade: He charged the Coalition of the Left and Progress had sold out its leftist tenets by collaborating in a right-wing plot . . . .* [wsj\_2431]

## 5.6 Newly Observed Ambiguities in Intra-sentential Connectives

### 5.6.1 Ambiguity of Concession Markers: Arg1-as-denier vs. Arg2-as-denier

While four of the subordinating conjunctions signalling CONCESSION (*although*, *even though*, *though* and *while*) could have been noticed as ambiguous in the PDTB-2 by virtue of having some tokens labelled CONTRA-EXPECTATION (ARG2-AS-DENIER in the PDTB-3 sense repertoire) and other tokens labelled EXPECTATION (ARG1-AS-DENIER in the PDTB-3 sense repertoire), it was not remarked on, even though in general, an explicit connective whose signalled sense is asymmetric will signal only one of the two directions. For example, when the adverbial *otherwise* conveys EXPANSION.EXCEPTION, it signals that *Arg1* is the exception, as in Ex. 245. In contrast, when the preposition *except* conveys EXPANSION.EXCEPTION, it signals that **Arg2** is the exception, as in Ex. 246.

(245) *Twenty-five years ago the poet Richard Wilbur modernized this 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.* [wsj\_0936]

(246) *Boston Co. officials declined to comment on Moodys action on the units 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.*

Our review and re-annotation of COMPARISON.CONTRAST and COMPARISON.CONCESSION tokens in the PDTB-2 (cf. Section 4.2.3) revealed an interesting ambiguity involving these four conjunctions. When *although*, *even though*, *though* and *while* occur in a preposed subordinate clause interpreted as COMPARISON.CONCESSION (394 tokens), the matrix clause (*Arg1*) is unambiguously interpreted as denying the expected causal relation conceded in the subordinate clause, **Arg2**, as in Ex. 247:

(247) The documents also said *that although the 64-year-old Mr. Cray has been working on the project for more than six years, the Cray-3 machine is at least another year away from a fully operational prototype.* [wsj\_0018]

In contrast, postposed subordinate clauses with *although*, *even though*, *though* and *while* (436 tokens) are ambiguous between interpreting the matrix clause (*Arg1*) as denying or cancelling the expected causal relation conceded in **Arg2**, as in Ex. 248, and the opposite, interpreting the subordinate clause (**Arg2**) as denying or cancelling the expected causal relation conceded in *Arg1*, as in Ex. 249:

- (248) The company’s research suggests *that its name recognition among most consumers remains unusually low, although (CONCESSION.ARG1-AS-DENIER)]* **its array of brands – including Maxwell House coffee, Jell-O, Cheez Whiz, and Miller beer – blanket supermarket shelves.** [wsj\_0326]
- (249) *Unemployment still is officially recorded at 16.5%, the highest rate in Europe,* although (CONCESSION.ARG2-AS-DENIER) **actual joblessness may be lower.** [wsj\_0456]

We have not yet carried out a detailed analysis of postposed subordinate CONCESSION tokens to determine whether anything more useful than *World Knowledge* will allow a system to predict accurately which way the ambiguity should be resolved.

### 5.6.2 Ambiguity of *So* and *so that*: Purpose vs. Result

While both *so* and *so that* were annotated as explicit connectives in the PDTB-2, the absence of PURPOSE in the sense hierarchy meant that the sense of all such tokens was given as RESULT. Adding PURPOSE to the PDTB-3 sense hierarchy suggested a review of the 31 tokens of *so that*, to determine which, if any, should be annotated as PURPOSE.ARG2-AS-GOAL. The review led to 21 tokens of *so that* being re-labelled as PURPOSE.ARG2-AS-GOAL (as in Ex. 250), while 10 retained the sense label RESULT (as in Ex. 251).

- (250) *Coach them in handling complaints* so that **they can resolve problems immediately** [wsj\_1504]
- (251) *The tremendous energy of the quake was dissipated by the distance* so that **most parts of the valley and the major cities suffered largely cosmetic damage – broken windows, falling brick and cornices, buckled asphalt or sidewalks** [wsj\_1870]

So *so that* must be seen as ambiguous between these two senses.

There are also contexts in which *so* on its own can be interpreted as either RESULT (as in Ex. 252) or PURPOSE.ARG2-AS-GOAL (as in Ex. 253, where *that* simply appears to have been dropped).

- (252) *Most murders are state crimes* so **any federal capital-punishment law probably would turn out to be more symbolism than substance** [wsj\_0426]
- (253) *Rep. Oakar, a Democrat from Cleveland, wants a \$6.9 million grant* so **Cleveland can build an 18-story Rock and Roll Hall of Fame** [wsj\_1847]

While all 114 tokens of sentence-initial *so* unambiguously convey RESULT, in clause-medial position, *so* can convey either of these two senses. As with the post-posed subordinate CONCESSION clauses mentioned in the previous subsection, we have not yet carried out a sufficiently detailed analysis that would allow a system to predict which way this RESULT/PURPOSE ambiguity should be resolved.

## 6 Inter-sentential Discourse Annotation

### 6.1 Inter-sentential relations missing from the PDTB-2

Fewer than 400 inter-sentential relations have been added to the PDTB-3 beyond those that were in the PDTB-2 (Figure 1), although the sense labels and/or argument spans of many more will have changed from the PDTB-2 to the PDTB-3 (cf. Section 3.5). Of the additional relations, about 60 should have been in the PDTB-2, as they were adjacent sentences within a paragraph that had simply not been annotated. The others were discovered as being missing in the course of other tasks.

### 6.2 Re-annotating PDTB-2 tokens due to changes in the sense-hierarchy

Of the  $\sim 35000$  tokens in the PDTB-2 with labels from the sense hierarchy (i.e., Explicit, Implicit or AltLex relations),  $\sim 25300$  were unchanged in the move from the PDTB-2 version of the sense hierarchy to the PDTB-3 version, while  $\sim 5600$  were automatically mapped 1:1 from their PDTB-2 sense label to their PDTB-3 sense label (e.g. mapping EXPANSION.RESTATEMENT.EQUIVALENCE to EXPANSION.EQUIVALENCE) and another  $\sim 4000$  automatically mapped from several PDTB-2 sense labels to a single PDTB-3 sense label. Sometimes, additional manual review was required: When all level-3 subtypes of CONTINGENCY.CONDITION were mapped to the PDTB-3 level-2 sense CONTINGENCY.CONDITION, manual review was required to decide which level-3 sense to assign to the token: ARG1-AS-COND, ARG2-AS-COND. Another  $\sim 600$  tokens had senses that were eliminated in going to the PDTB-3 hierarchy (e.g., EXPANSION.ALTERNATIVE, EXPANSION.RESTATEMENT and their subtypes), that then had to be sense-relabelled manually.

In addition, the addition of new senses such as COMPARISON.SIMILARITY and CONTINGENCY.PURPOSE meant that other tokens had to be reviewed to see if their label should be changed to one of these new ones.

## 7 Consistency Checking

The standard practice of assessing Inter-annotator Agreement (IAA), while important in corpus annotation, does not eliminate all problems. In particular, assessing IAA does not guarantee that the same phenomenon has been annotated consistently, in the same way throughout the corpus. This can be especially problematic in a large corpus, in which many different, and often unexpected, semantic phenomena are regularly encountered and where annotation takes many months or even years to complete. The PDTB-3 has involved nearly three years of annotation and re-annotation, although this was partly a problem of student annotators withdrawing from annotation during the first year and having to engage and train additional annotators, who also moved on to other things before the annotation was complete.) This has required us to impose rigorous new semantic consistency checks over parts of the corpus and then over the corpus as a whole, in order to discover anomalies and possible inconsistencies. These were then reviewed and corrected if they didn't just turn out to be some low-frequency examples that were correctly, if oddly, annotated.

Consistency checking of Explicit discourse relations required identifying all the Explicit token in the corpus, ordering them by their connective and their senses and getting frequency counts for all pairings. The same needed to be done for all AltLex discourse relations. All implicit discourse relations in the

corpus need to be assembled and ordered by their set of senses, again with frequency counts for all pairings. This information was then used in carrying out the following checks, often for each sub-task in preparing the corpus and then again at the end, as the same tokens may have ended up being annotated more than once.

- Explicit connectives and their senses: For each explicit connective, what senses has it been said to express? Review any pairings that seem anomalous. If an explicit connective has been labelled as simultaneously conveying multiple senses, do any of the sets seem anomalous? Should any of the senses be associated with an implicit connective, reflecting a sense that is inferred from material in the argument spans, rather than from the connective itself.
- Explicit connectives and alternative lexicalizations: For each explicit connective, has it ever been labelled as an AltLex? Is that inconsistent, or is it also a valid AltLex as an adjective or adjective modifier (Section 3.2)?
- Alternative lexicalizations: For each AltLex, does it always appear with a similar text span as its alternative lexicalization, or does it vary between tokens? Is the latter an inconsistency that needs to be corrected, or does the variation actually correspond to different alternative lexicalizations?
- Implicit connectives: When a token has been annotated with multiple implicit connectives, each with its own sense, does the pair of senses commonly occur together or is it a potential inconsistency?
- Linked explicit and implicit tokens: When an explicit relation shares its arguments with an implicit relation, each with its own sense, does the pair of senses commonly occur together or is it a potential anomaly?
- Intra-sentential discourse relations with overlapping arguments that reflect different, incompatible interpretations of a sentence. Here is an example.

(254) When angered, he cursed so forcefully that his face reddened and his pale-blue eyes narrowed into tiny slits. [wsj\_1986]

Ex. 254 contains two intra-sentential discourse relations to be annotated, one associated with the explicit conjunction *and*, and the other associated with the construction *so forcefully that*. If they are annotated as in Ex. 255, the annotation would be inconsistent because Ex. 255a interprets his eyes narrowing into slits as being a result of his being angered but **not** as a result of the force of his cursing, while Ex. 255b interprets his eyes narrowing into slits as being a result of both being angered and cursing forcefully.

(255) a. When angered, *he cursed so forcefully that his face reddened* and **his pale-blue eyes narrowed into tiny slits.**

b. When angered, *he cursed so forcefully* that **his face reddened and his pale-blue eyes narrowed into tiny slits**

## 8 Implementation

### 8.1 Format of Relation Tokens

PDTB-3 uses a simple data format in which each relation token is stored as a pipe-delimited row of text. The more complex file format of PDTB-2, which was designed to work with Lex/Yacc, has been

abandoned in favor of this simpler format which is accessible to a wider range of users. For example, a PDTB-3 annotation file can be easily imported into a spreadsheet or parsed into its component fields using a few lines of a scripting language such as Python or Perl. The annotation tool used for PDTB-3 — the PDTB Annotator (Lee et al., 2016) — makes use of this newer data format.

To maintain backward compatibility, data fields from PDTB-2 have been preserved in the PDTB-3 format, including ones deprecated in PDTB-3. For example, while attribution features are not annotated in PDTB-3, these fields are retained. A tool for converting PDTB-2 annotation files into the PDTB-3 format is provided in the release. The fields in the PDTB-3 data format are shown in Tables 2–3.

Index	Field Name	Description
0	Relation Type	Explicit, Implicit, AltLex, AltLexC, Hypophora, EntRel, NoRel
1	Conn SpanList	SpanList of the Explicit Connective or the AltLex/AltLexC selection
2	Conn Src	Connective’s Source
3	Conn Type	Connective’s Type
4	Conn Pol	Connective’s Polarity
5	Conn Det	Connective’s Determinacy
6	Conn Feat SpanList	Connective’s Feature SpanList
7	Conn1	Explicit Connective Head / First Implicit Connective
8	SClass1A	First Semantic Class of the First Connective
9	SClass1B	Second Semantic Class of the First Connective
10	Conn2	Second Implicit Connective
11	SClass2A	First Semantic Class of the Second Connective
12	SClass2B	Second Semantic Class of the Second Connective
13	Sup1 SpanList	SpanList of the First Argument’s Supplement
14	Arg1 SpanList	SpanList of the First Argument
15	Arg1 Src	First Argument’s Source
16	Arg1 Type	First Argument’s Type

Table 2: PDTB-3 Data Fields (part 1)

Tokens are ordered in a file by the start point of the Offset (field 31).

## 8.2 Provenance

The Provenance field (field 32) contains one of three tags:

- PDTB3
- PDTB2::wsj\_XXXX::{PDTB2 offset}::SAME
- PDTB2::wsj\_XXXX::{PDTB2 offset}::CHANGED

The tag PDTB-3 indicates that the token is new. The other two indicate a unique corresponding token in file wsj\_XXXX of the PDTB-2 corpus, identified by the token’s PDTB2 unique id, which is the offset of its PDTB-2 token. The offset is either the Conn SpanList (for a PDTB2 Explicit or AltLex token) or the Inference site (for a PDTB-2 Implicit, EntRel or NoRel token). SAME means the PDTB3 token is unchanged from its PDTB-2 correspondent for Senses, argument SpanLists, Explicit connective or AltLex SpanLists, or Implicit connective strings. CHANGED means one of those things have been modified. Attribution spans and features have been ignored in tracking changes.

17	Arg1 Pol	First Argument's Polarity
18	Arg1 Det	First Argument's Determinacy
19	Arg1 Feat SpanList	SpanList of the First Argument's Feature
20	Arg2 SpanList	SpanList of the Second Argument
21	Arg2 Src	Second Argument's Source
22	Arg2 Type	Second Argument's Type
23	Arg2 Pol	Second Argument's Polarity
24	Arg2 Det	Second Argument's Determinacy
25	Arg2 Feat SpanList	SpanList of the Second Argument's Feature
26	Sup2 SpanList	SpanList of the Second Argument's Supplement
27	Adju Reason	The Adjudication Reason
28	Adju Disagr	The type of the Adjudication disagreement
29	PB Role	The PropBank role of the PropBank verb
30	PB Verb	The PropBank verb of the main clause of this relation
31	Offset	The Conn SpanList of Explicit/AltLex/AltLexC tokens or the start point of the Arg2 of an Implicit/Hypophora/EntRel/NoRel tokens
32	Provenance	Indicates whether the token is a new PDTB3 token or has a corresponding PDTB2 token (see 8.2)
33	Link	The link id of the token

Table 3: PDTB-3 Data Fields (part 2)

Note that the use of the offset to uniquely identify a token in PDTB-2 is no longer applicable in PDTB-3, where two non-Explicit tokens could potentially share the same inference site. The offset in PDTB-3 is used for ordering tokens but not for uniquely identifying them. A PDTB-3 token can be uniquely identified using a combination of the Relation Type, Arg1 SpanList and Arg2 SpanList.

### 8.3 Annotation Process

Annotation of the PDTB-3 was carried out in basically independent *tasks*, each focused on a particular objective such as annotation of particular syntactic structures (infinitival clauses, free adjuncts, conjoined verb phrases, conjoined clauses, etc.), annotation a particular relation type (Hypophora), annotation of inter-sentential tokens missing from the PDTB-2, review of Contrast vs Concession tokens from the PDTB-2 to reflect a more precise distinction, etc. As such, a task might introduce new hitherto unannotated tokens or re-annotate, where necessary, subsets of the PDTB-2 corpus. In all cases, an annotation task was carried out independently of all other tasks, with annotators working on a set of candidate tokens pre-identified by criteria relevant to the task at hand and with no reference to other tokens external to the task.

The final corpus for PDTB-3 was then put together by combining the annotated tokens from the various tasks along with tokens from the original PDTB-2 corpus. This had to be done with care, to ensure that the most up-to-date tokens were included in the corpus, and that clashes between tokens annotated in different tasks were resolved.

### 8.4 Overlapping Tokens

A particular challenge that arose when combining tokens for the final corpus was to identify *overlaps* between tokens within the same annotation file and ascertain which reflected distinct relations and which were conflicting variants of the same relation, only one of which should be retained. Both were possible,

given that overlaps could occur between tokens coming from different tasks or between a token from some task and an original PDTB-2 token. More precisely, two tokens from the same annotation file overlap when they share one or more of the following characteristics:

1. They have identical or overlapping *overt* SpanLists, for Explicit, AltLex and AltLexC relations;
2. They share the same inference site, where an inference site is either the starting offset of Arg2 of a non-Explicit token, or the starting offset of either the Arg2 SpanList or the Explicit connective SpanList of an Explicit token, whichever offset is smaller;
3. They have overlapping Arg1 and Arg2 spans.

Two overlapping tokens are deemed to be in conflict if they turn out to be alternative annotations of a single discourse relation. In such cases, a *clash* has occurred and the more appropriate annotation needs to be selected, either automatically according to some criteria or through manual adjudication. But not all overlaps correspond to clashes, since the PDTB-3 also allows overlapping relations to be linked (cf. Section 8.5) when they are distinct and there is evidence for both (including inferential evidence). Hence, in reviewing a pair of overlapping tokens, a decision needed to be made as to whether:

- the tokens clash, in which case one of the tokens had to be discarded;
- the tokens correspond to different relations that should be linked according to the PDTB-3 criteria for linking, in which case both tokens were accepted.

Many of the overlaps could be resolved automatically. In the tasks associated with conjoined VPs (Section 5.4) and conjoined clauses (Section 5.5), links between tokens had been established during adjudication and explicitly noted in their annotation, so overlapping tokens marked with these links were accepted into the final corpus. For review tasks, correspondences between tokens from the PDTB-2 and the output of the review were recorded, and any reviewed token which differed from its original PDTB-2 correspondent was accepted into the corpus and the original discarded. Where there were no changes, only one copy (from the PDTB-2) was retained.

In resolving the remaining overlaps, the following rules were applied:

1. Identical tokens were merged;
2. Overlaps of type (1) — identical or overlapping overt SpanLists, as defined above — were classified as a clash, since there should be no multiple relations annotated on a single overt SpanList (e.g. an Explicit connective like "because" in a text should never be annotated twice, i.e. as two tokens with different senses and/or Arg1 and Arg2 spans). If needed, this token can be annotated with multiple senses.
3. Overlaps of types (2) and (3) can be either a clash or a link.

To resolve clashes, the following rules were applied:

- Any clash in which one token was of type Explicit, Implicit, AltLex, AltLexC or Hypophora and the other token was of type EntRel or NoRel was resolved in favor of the former;
- Any clash in which one token was of type Explicit, AltLex or AltLexC and the other token was of type Implicit and where both tokens carried the same sense, was resolved in favour of the former;

- Any clash in which one token was of type Explicit, AltLex or AltLexC and the other token was of type Implicit with a sense of Expansion.Conjunction was resolved in favor of the former;
- Any clash in which both tokens were of type Implicit, with one having a non Expansion.Conjunction sense and the other, an Expansion.Conjunction sense, was resolved in favor of the former;
- Any clash in which one token was an original PDTB-2 token and the other was from a PDTB-3 task was resolved in favor of the latter.

All overlaps that remained unresolved after applying the above rules were marked for manual adjudication.

## 8.5 Format of Multiple Relations

Multiple relations can hold between a pair of spans in four circumstances:

- where a single explicit connective has multiple senses (e.g., *since* conveying both REASON and PRECEDENCE,
- where in the absence of any explicit connective, annotators infer more than one sense as holding between a pair of spans,
- where there are multiple explicit connectives between the same spans,
- where an explicit connective or AltLex or AltLexC conveys one relation between a pair of spans, but annotators infer a separate sense as well.

Each case is represented in a different way.

In the case of a single explicit connective with multiple senses, the senses are annotated as SClass1A (field 8) and SClass1B (field 9) on the same token – e.g.

(256) *I cannot recall any disorder in currency markets* since (SUCCESSION,REASON) **the 1974 guidelines were adopted.** [wsj\_0779]

```
Explicit|4940..4945| || || |since|Temporal.Asynchronous.Succession|Contingency.Cause.Reason| || || |4891..4939| ||
|| |4946..4978| || || || || |ARGM-TMP|recall|4940..4945|PDBT3|
```

When more than one sense has been inferred as holding between two spans, the senses are annotated as SClass1A (field 8) and SClass2A (field 11) on the same token, with the implicit connectives expressing these senses annotated as Conn1 (field 7) and Conn2 (field 10). (The field structure is given in Section 8.1.)

(257) *... there's also a price to pay* (Implicit=REASON,ARG2-AS-DETAIL) **A diversified portfolio always underperforms an undiversified portfolio during those times when the investment in the undiversified portfolio is truly hot** [wsj\_1453]

```
Implicit| || || || || |because|Contingency.Cause.Reason| || |specifically|Expansion.Level-of-detail.Arg2-as-
detail| || |2566..2593| || || || |2595..2749| || || || || || |2595|PDTB2::wsj_1453::2595..2749::CHANGED|
```

Where multiple explicit connectives occur between the same spans, each is annotated on a separate token, with the tokens then *linked* with the same index value in their link fields (field 33) – e.g.

(258) *Small businesses say a recent trend is like a dream come true: more-affordable rates for employee-health insurance, initially at least* **But** then (PRECEDENCE) **they wake up to a nightmare** [wsj\_0518]

```
Explicit|149..153| ||| |then|Temporal.Asynchronous.Precedence| |||
| ||9..143| ||| |145..148;154..181| ||| | ||| |149..153|PDTB2::wsj_0518::149..153::SAME|LINK1
```

(259) *Small businesses say a recent trend is like a dream come true: more-affordable rates for employee-health insurance, initially at least* But (ARG2-AS-DENIER) *argii[then they wake up to a nightmare]* [wsj\_0518]

```
Explicit|145..148| ||| |but|Comparison.Concession.Arg2-as-
denier| ||| |9..143| ||| |149..181| ||| | ||| |145..148|PDTB2::wsj_0518::145..148::SAME|LINK1
```

Finally, where annotators infer a sense that is separate from that conveyed by an explicit connective or AltLex or AltLexC, the implicit relation is annotated on a separate token from the explicit relation, with the tokens then *linked* with the same index value in their link fields (field 33). (As noted in Section 8.1, tokens are recorded in order of the starting position of Arg2 (field 20).)

(260) We've got to *get out of the Detroit mentality* and (EXPANSION.CONJUNCTION) Implicit=instead (EXPANSION.SUBSTITUTION.ARG2-AS-SUBST) **be part of the world mentality,**” declares Charles M. Jordan, GM’s vice president for design ... [wsj\_0956]

```
Implicit| ||| | ||| |instead|Expansion.Substitution.Arg2-as-subst| ||| |5914..5946| ||| |5947..5981| ||| | ||| |5947
|PDTB3|LINK4
```

```
Explicit|5947..5950| ||| |and|Expansion.Conjunction| ||| |5914..5946| ||| |5951..5981| ||| | ||| |5947..5950
|PDTB3|LINK4
```

What cannot be annotated directly are cases where an explicit connective is “parasitic” on an implicit relation, as in

(261) *Despite the current outcry over stock market volatility, few people expect stock futures to disappear.* **For one thing, the Chicago futures exchanges have political and financial clout – including many friends in Congress** [wsj\_0742]

While *for one thing* indicates that **Arg2** is one REASON for the claim in *Arg1*, the annotation system does not permit saying this directly. The only thing that can be done is to annotate both ARG2-AS-INSTANCE and REASON on the altlex token for *for one thing* – e.g.

```
AltLex|4184..4197| ||| | ||| |Contingency.Cause.Reason|Expansion.Instantiation.Arg2-as-
instance| ||| |4081..4182| ||| |4184..4301| ||| | ||| |4184..4197|PDTB3|
```

## References

- Jesse Dunietz, Lori Levin, and Jaime Carbonell. 2017. Automatically tagging constructions of causation and their slot-fillers. *Transactions of the Association for Computational Linguistics (ACL)* 5:117–133.
- Alan Lee, Rashmi Prasad, Bonnie Webber, and Aravind K. Joshi. 2016. Annotating discourse relations with the PDTB annotator. In *Proceedings, 26th Int’l Conference on Computational Linguistics (COLING)*. pages 121–125.

- Mitchell Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn TreeBank. *Computational Linguistics* 19(2):313–330.
- Francois Mouret. 2006. A phrase structure approach to argument cluster coordination. In Stephan Müller, editor, *Proceedings, 13th International Conference on Head-Driven Phrase Structure Grammar*. pages 247–267.
- Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational linguistics* 31(1):71–106.
- Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings, 6th International Conference on Language Resources and Evaluation*. Marrakech, Morocco, pages 2961–2968.
- Rashmi Prasad, Katherine Forbes-Riley, and Alan Lee. 2017. Towards full text shallow discourse relation annotation: Experiments with cross-paragraph implicit relations in the PDTB. In *Proc, 18<sup>th</sup> Annual SIGdial Meeting on Discourse and Dialogue*. pages 7–16.
- Rashmi Prasad, Bonnie Webber, Alan Lee, Sameer Pradhan, and Aravind Joshi. 2015. Bridging sentential and discourse-level semantics through clausal adjuncts. In *Proceedings of the First Workshop on Linking Computational Models of Lexical, Sentential and Discourse-level Semantics*. pages 64–69.
- Tatjana Scheffler and Manfred Stede. 2016. Adding semantic relations to a large-coverage connective lexicon of German. In *Proceedings, 10th International Conference on Language Resources and Evaluation, LREC-16*. Potorož, Slovenia, pages 1008–1013.
- Mark Steedman. 1989. Constituency and coordination in combinatory grammar. In M. Baltin and A. Kroch, editors, *Alternative Conceptions of Phrase Structure*, University of Chicago Press, pages 201–231.
- Mark Steedman. 2000. *The Syntactic Process*. MIT Press.
- Bonnie Webber. 2013. What excludes an alternative in coherence relations? In *Proceedings of the 10th International Conference on Computational Semantics*. Potsdam, Germany, pages 276–287.
- Nianwen Xue, Hwee Tou Ng, Sameer Pradhan, Rashmi Prasad, Christopher Bryant, and Attapol Rutherford. 2015. The CoNLL-2015 shared task on shallow discourse parsing. In *Proceedings of the CoNLL-15 shared task*. Beijing, China, pages 1–16.
- Nianwen Xue, Hwee Tou Ng, Attapol Rutherford, Bonnie Webber, Chuan Wang, and Hongmin Wang. 2016. The CoNLL-2016 shared task on multilingual shallow discourse parsing. In *Proceedings of the CoNLL-16 shared task*. Berlin, Germany, pages 1–19.
- Deniz Zeyrek, Amalia Mendes, and Murathan Kurfali. 2018. Multi-lingual extension of PDTB-style annotation: The case of TED multi-lingual discourse bank. In *Proceedings, 11th International Conference on Language Resources and Evaluation, LREC-18*. Miyazaki, Japan.

# Appendix

## A Explicit Connectives and their Senses

This section provides the distribution of the 24,240 explicit connectives and their senses. Discontinuous connectives are indicated with a ”+” symbol between their parts.

about	2	Contingency.Cause.Reason
accordingly	5	Contingency.Cause.Result
additionally	7	Expansion.Conjunction
after	56 533	Temporal.Asynchronous.Succession Contingency.Cause.Reason Temporal.Asynchronous.Succession
afterward	6	Temporal.Asynchronous.Precedence
afterwards	5	Temporal.Asynchronous.Precedence
albeit	1	Comparison.Concession.Arg2-as-denier
along with	2	Expansion.Conjunction
also	1736 1	Expansion.Conjunction Temporal.Synchronous Expansion.Conjunction
alternatively	4 2	Expansion.Disjunction Expansion.Substitution.Arg2-as-subst
although	206 105 14 2 1	Comparison.Concession.Arg1-as-denier Comparison.Concession.Arg2-as-denier Comparison.Contrast Expansion.Exception.Arg2-as-excpt Temporal.Synchronous Comparison.Contrast
and	2 16 1 4 12 5 22 128 6189 4 19	Comparison.Concession+SpeechAct.Arg2-as-denier+SpeechAct Comparison.Contrast Contingency.Cause+SpeechAct.Result+SpeechAct Contingency.Cause.Reason Contingency.Cause.Result Expansion.Conjunction Contingency.Cause.Result Contingency.Condition.Arg1-as-cond Contingency.Purpose.Arg2-as-goal Expansion.Conjunction Expansion.Level-of-detail.Arg2-as-detail Expansion.Manner.Arg2-as-manner
and then	1	Expansion.Disjunction
and/or	2	Expansion.Conjunction Expansion.Disjunction

as	7	Comparison.Concession.Arg1-as-denier
	5	Comparison.Contrast
	65	Comparison.Similarity
	3	Contingency.Cause+Belief.Reason+Belief
	180	Contingency.Cause.Reason
	4	Expansion.Instantiation.Arg2-as-instance
	8	Expansion.Level-of-detail.Arg1-as-detail
	3	Expansion.Manner.Arg2-as-manner
	23	Temporal.Asynchronous.Succession Contingency.Cause.Reason
	3	Temporal.Asynchronous.Succession
	2	Temporal.Synchronous Comparison.Contrast
	1	Temporal.Synchronous Comparison.Similarity
	5	Temporal.Synchronous Contingency.Cause+Belief.Reason+Belief
	146	Temporal.Synchronous Contingency.Cause.Reason
383	Temporal.Synchronous	
as a result	78	Contingency.Cause.Result
as an alternative	2	Expansion.Disjunction
as if	4	Comparison.Concession.Arg2-as-denier
	10	Comparison.Similarity Expansion.Manner.Arg2-as-manner
	1	Expansion.Instantiation.Arg1-as-instance
	1	Expansion.Manner.Arg2-as-manner
as long as	13	Contingency.Condition.Arg2-as-cond
	7	Temporal.Synchronous Contingency.Condition.Arg2-as-cond
	4	Temporal.Synchronous
as much as	2	Comparison.Concession.Arg1-as-denier
	1	Expansion.Conjunction
	1	Expansion.Substitution.Arg2-as-subst
as soon as	11	Temporal.Asynchronous.Succession
	9	Temporal.Synchronous
as though	2	Comparison.Similarity Expansion.Manner.Arg2-as-manner
	1	Comparison.Similarity
	2	Expansion.Level-of-detail.Arg2-as-detail
as well	6	Comparison.Similarity
	12	Expansion.Conjunction
as well as	7	Expansion.Conjunction
at the same time	65	Temporal.Synchronous
because	2	Contingency.Cause+Belief.Reason+Belief
	833	Contingency.Cause.Reason
	2	Contingency.Condition+SpeechAct
because of	12	Contingency.Cause.Reason
before	309	Temporal.Asynchronous.Precedence
	2	Temporal.Asynchronous.Succession
before and after	1	Temporal.Asynchronous.Precedence Temporal.Asynchronous.Succession
besides	19	Expansion.Conjunction
beyond	1	Expansion.Conjunction

both+and	6	Expansion.Conjunction
but	2	Comparison.Concession+SpeechAct.Arg2-as-denier+SpeechAct
	3063	Comparison.Concession.Arg2-as-denier
	618	Comparison.Contrast
	1	Contingency.Cause+SpeechAct.Reason+SpeechAct
	1	Contingency.Cause.Reason Comparison.Concession.Arg2-as-denier
	42	Expansion.Conjunction
	3	Expansion.Exception.Arg2-as-excpt
	1	Temporal.Synchronous Comparison.Contrast
but also	1	Expansion.Conjunction
but then	3	Comparison.Concession.Arg2-as-denier
but then again	1	Comparison.Concession.Arg2-as-denier
by	1	Comparison.Concession.Arg1-as-denier
	3	Contingency.Cause+Belief.Reason+Belief Expansion.Manner.Arg2-as-manner
	95	Contingency.Cause.Reason Expansion.Manner.Arg2-as-manner
	10	Contingency.Cause.Reason
	32	Contingency.Condition.Arg2-as-cond Expansion.Manner.Arg2-as-manner
	2	Contingency.Condition.Arg2-as-cond
	115	Contingency.Purpose.Arg1-as-goal Expansion.Manner.Arg2-as-manner
	2	Expansion.Level-of-detail.Arg2-as-detail
174	Expansion.Manner.Arg2-as-manner	
1	Temporal.Synchronous	
by comparison	11	Comparison.Contrast
by contrast	28	Comparison.Contrast
by then	1	Temporal.Asynchronous.Succession Contingency.Cause.Reason
	6	Temporal.Asynchronous.Succession
consequently	10	Contingency.Cause.Result
conversely	2	Comparison.Contrast
depending on	3	Contingency.Condition.Arg2-as-cond
depending upon	1	Contingency.Condition.Arg2-as-cond
despite	9	Comparison.Concession.Arg1-as-denier
due to	1	Contingency.Cause.Reason
earlier	15	Temporal.Asynchronous.Succession
either+or	2	Contingency.Negative-condition.Arg1-as-negCond
	36	Expansion.Disjunction
else	1	Contingency.Negative-condition.Arg1-as-negCond
even after	8	Temporal.Asynchronous.Succession Comparison.Concession.Arg1-as-denier
even as	2	Comparison.Concession.Arg1-as-denier
	11	Temporal.Synchronous Comparison.Concession.Arg1-as-denier
even before	14	Temporal.Asynchronous.Precedence Comparison.Concession.Arg1-as-denier
even before then	1	Temporal.Asynchronous.Succession Comparison.Concession.Arg2-as-denier
even if	87	Comparison.Concession.Arg1-as-denier
even so	9	Comparison.Concession.Arg2-as-denier
even then	2	Temporal.Asynchronous.Precedence Comparison.Concession.Arg2-as-denier
even though	69	Comparison.Concession.Arg1-as-denier
	26	Comparison.Concession.Arg2-as-denier

even when	8	Comparison.Concession.Arg1-as-denier
	4	Temporal.Asynchronous.Succession Comparison.Concession.Arg1-as-denier
	2	Temporal.Synchronous Comparison.Concession.Arg1-as-denier
even while	3	Temporal.Synchronous Comparison.Concession.Arg1-as-denier
even with	2	Comparison.Concession.Arg1-as-denier
except	12	Expansion.Exception.Arg2-as-excpt
finally	18	Expansion.Conjunction
	1	Temporal.Asynchronous.Precedence Contingency.Cause.Result
	13	Temporal.Asynchronous.Precedence
for	1	Comparison.Concession.Arg1-as-denier
	34	Contingency.Cause.Reason
	1	Contingency.Cause.Result
	7	Contingency.Condition.Arg2-as-cond
	16	Contingency.Purpose.Arg2-as-goal
	1	Expansion.Level-of-detail.Arg2-as-detail
for example	200	Expansion.Instantiation.Arg2-as-instance
for instance	98	Expansion.Instantiation.Arg2-as-instance
from	2	Contingency.Cause+Belief.Reason+Belief
	1	Contingency.Cause.Reason Contingency.Condition.Arg2-as-cond
	1	Contingency.Cause.Reason Expansion.Manner.Arg2-as-manner
	2	Contingency.Cause.Reason
	1	Expansion.Substitution.Arg1-as-subst
further	7	Expansion.Conjunction
furthermore	12	Expansion.Conjunction
given	3	Contingency.Cause+Belief.Reason+Belief
	6	Contingency.Cause.Reason
hence	5	Contingency.Cause.Result
however	5	Comparison.Concession.Arg1-as-denier
	390	Comparison.Concession.Arg2-as-denier
	95	Comparison.Contrast
	1	Temporal.Synchronous Comparison.Contrast
if	1	Comparison.Concession+SpeechAct.Arg2-as-denier+SpeechAct
	6	Comparison.Concession.Arg1-as-denier
	3	Comparison.Concession.Arg2-as-denier
	2	Comparison.Contrast
	56	Contingency.Condition+SpeechAct
	1	Contingency.Condition.Arg2-as-cond Expansion.Level-of-detail.Arg2-as-detail
if and when	1084	Contingency.Condition.Arg2-as-cond
	1	Temporal.Synchronous Contingency.Condition.Arg2-as-cond
	2	Contingency.Condition.Arg2-as-cond
	1	Temporal.Synchronous Contingency.Condition.Arg2-as-cond
if only	1	Comparison.Concession.Arg2-as-denier
	4	Contingency.Condition.Arg2-as-cond
	1	Contingency.Purpose.Arg2-as-goal
if+then	1	Contingency.Condition+SpeechAct
	37	Contingency.Condition.Arg2-as-cond

in	1	Contingency.Cause+Belief.Reason+Belief
	3	Contingency.Cause.Reason Expansion.Manner.Arg2-as-manner
	1	Contingency.Cause.Reason
	2	Contingency.Condition.Arg2-as-cond Expansion.Manner.Arg2-as-manner
	8	Contingency.Condition.Arg2-as-cond
	2	Contingency.Purpose.Arg2-as-goal
	1	Expansion.Instantiation.Arg1-as-instance
	17	Expansion.Level-of-detail.Arg1-as-detail
	2	Expansion.Level-of-detail.Arg2-as-detail
	13	Expansion.Manner.Arg2-as-manner
	1	Temporal.Synchronous Contingency.Purpose.Arg2-as-goal
	1	Temporal.Synchronous Expansion.Level-of-detail.Arg1-as-detail
	18	Temporal.Synchronous
in addition	165	Expansion.Conjunction
in any case	3	Comparison.Concession.Arg2-as-denier
in case	6	Contingency.Condition.Arg2-as-cond
in contrast	12	Comparison.Contrast
in fact	4	Comparison.Concession.Arg2-as-denier
	7	Comparison.Contrast
	36	Expansion.Conjunction
	3	Expansion.Instantiation.Arg2-as-instance
	1	Expansion.Level-of-detail.Arg1-as-detail
	34	Expansion.Level-of-detail.Arg2-as-detail
in order	4	Contingency.Condition.Arg2-as-cond
	51	Contingency.Purpose.Arg2-as-goal
in other words	17	Expansion.Equivalence
in particular	6	Expansion.Instantiation.Arg2-as-instance
	9	Expansion.Level-of-detail.Arg2-as-detail
in short	4	Expansion.Level-of-detail.Arg1-as-detail
in sum	2	Expansion.Level-of-detail.Arg1-as-detail
in that	1	Expansion.Level-of-detail.Arg2-as-detail
in the end	1	Comparison.Concession.Arg2-as-denier
	1	Comparison.Contrast
	2	Contingency.Cause.Result
	1	Expansion.Conjunction
	2	Expansion.Level-of-detail.Arg1-as-detail
	1	Expansion.Level-of-detail.Arg2-as-detail
3	Temporal.Asynchronous.Precedence	
in the meantime	2	Temporal.Asynchronous.Succession
	1	Temporal.Synchronous Comparison.Contrast
	11	Temporal.Synchronous
in the meanwhile	1	Temporal.Synchronous

indeed	1	Comparison.Concession.Arg2-as-denier
	4	Contingency.Cause+Belief.Reason+Belief
	1	Contingency.Cause.Reason
	1	Contingency.Cause.Result Expansion.Conjunction
	54	Expansion.Conjunction
	2	Expansion.Equivalence
	2	Expansion.Instantiation.Arg2-as-instance
	1	Expansion.Level-of-detail.Arg1-as-detail
insofar as	37	Expansion.Level-of-detail.Arg2-as-detail
	1	Contingency.Cause.Reason
instead	1	Expansion.Level-of-detail.Arg2-as-detail
	112	Expansion.Substitution.Arg2-as-subst
instead of	43	Expansion.Substitution.Arg1-as-subst
later	92	Temporal.Asynchronous.Precedence
later on	2	Temporal.Asynchronous.Precedence
lest	2	Contingency.Negative-condition.Arg1-as-negCond
like	1	Comparison.Contrast
	3	Comparison.Similarity
	1	Expansion.Instantiation.Arg2-as-instance
likewise	8	Expansion.Conjunction
meantime	2	Temporal.Synchronous
meanwhile	2	Comparison.Concession.Arg2-as-denier
	7	Comparison.Contrast
	27	Expansion.Conjunction
	1	Temporal.Synchronous Comparison.Concession.Arg2-as-denier
	33	Temporal.Synchronous Comparison.Contrast
	2	Temporal.Synchronous Comparison.Similarity
120	Temporal.Synchronous	
more accurately	1	Expansion.Substitution.Arg2-as-subst
moreover	103	Expansion.Conjunction
much less	3	Expansion.Conjunction
neither+nor	1	Comparison.Contrast
	2	Expansion.Conjunction
nevertheless	32	Comparison.Concession.Arg2-as-denier
	12	Comparison.Contrast
next	1	Temporal.Asynchronous.Precedence Expansion.Conjunction
	4	Temporal.Asynchronous.Precedence
no matter	8	Comparison.Concession.Arg1-as-denier
nonetheless	25	Comparison.Concession.Arg2-as-denier
	2	Comparison.Contrast

nor	1	Comparison.Cession.Arg2-as-denier
	31	Expansion.Conjunction
	1	Expansion.Disjunction
not just+but	1	Expansion.Conjunction
not just+but+also	1	Expansion.Conjunction
not only	5	Expansion.Conjunction
not only+also	1	Expansion.Conjunction
not only because of	1	Contingency.Cause.Reason
not only+but	1	Comparison.Cession.Arg2-as-denier
	18	Expansion.Conjunction
not only+but also	1	Comparison.Contrast
	9	Expansion.Conjunction
not so much as	1	Expansion.Substitution.Arg2-as-subst
now that	10	Contingency.Cause.Reason
	1	Temporal.Asynchronous.Precedence Contingency.Cause.Reason
	4	Temporal.Asynchronous.Succession Contingency.Cause.Reason
	4	Temporal.Synchronous Contingency.Cause.Reason
	2	Temporal.Synchronous
on	1	Contingency.Cause.Reason
on the contrary	4	Comparison.Contrast
on the one hand+on the other	1	Comparison.Contrast
on the one hand+on the other hand	1	Comparison.Cession.Arg2-as-denier
	1	Comparison.Contrast
on the other hand	4	Comparison.Cession.Arg2-as-denier
	32	Comparison.Contrast
once	4	Contingency.Condition.Arg2-as-cond
	7	Temporal.Asynchronous.Succession Contingency.Cause.Reason
	3	Temporal.Asynchronous.Succession Contingency.Condition.Arg2-as-cond
	70	Temporal.Asynchronous.Succession
only	2	Comparison.Cession.Arg2-as-denier
	1	Comparison.Contrast
	3	Expansion.Exception.Arg2-as-excpt
	1	Expansion.Level-of-detail.Arg2-as-detail
only if	13	Contingency.Condition.Arg2-as-cond
or	11	Comparison.Cession+SpeechAct.Arg2-as-denier+SpeechAct
	1	Comparison.Cession.Arg2-as-denier
	2	Contingency.Condition+SpeechAct
	7	Contingency.Negative-condition.Arg1-as-negCond
	71	Expansion.Conjunction
	258	Expansion.Disjunction
or otherwise	6	Expansion.Equivalence
	2	Expansion.Disjunction
otherwise	4	Contingency.Negative-condition.Arg1-as-negCond
	15	Expansion.Exception.Arg1-as-excpt
plus	1	Expansion.Conjunction
previously	1	Temporal.Asynchronous.Succession Comparison.Contrast
	53	Temporal.Asynchronous.Succession

rather	17	Expansion.Substitution.Arg2-as-subst
rather than	40	Expansion.Substitution.Arg1-as-subst
regardless	2	Comparison.Concession.Arg2-as-denier
regardless of	6	Comparison.Concession.Arg1-as-denier
separately	72 2	Expansion.Conjunction Temporal.Synchronous Expansion.Conjunction
similarly	18	Comparison.Similarity
simultaneously	6	Temporal.Synchronous
since	96 10 11 83	Contingency.Cause.Reason Temporal.Asynchronous.Precedence Temporal.Asynchronous.Succession Contingency.Cause.Reason Temporal.Asynchronous.Succession
since before	1	Temporal.Asynchronous.Succession
so	1 222 44	Contingency.Cause+Belief.Result+Belief Contingency.Cause.Result Contingency.Purpose.Arg2-as-goal
so as	3	Contingency.Purpose.Arg2-as-goal
so long as	4	Contingency.Condition.Arg2-as-cond
so much as	1	Expansion.Substitution.Arg2-as-subst
so that	10 21	Contingency.Cause.Result Contingency.Purpose.Arg2-as-goal
specifically	10	Expansion.Level-of-detail.Arg2-as-detail
still	115 75 2 1	Comparison.Concession.Arg2-as-denier Comparison.Contrast Temporal.Asynchronous.Precedence Temporal.Synchronous
subsequently	3	Temporal.Asynchronous.Precedence
such as	2	Expansion.Instantiation.Arg2-as-instance
that is	2 2	Expansion.Equivalence Expansion.Level-of-detail.Arg2-as-detail
then	1 7 1 11 5 310 4	Contingency.Cause.Result Expansion.Conjunction Contingency.Cause.Result Contingency.Condition.Arg1-as-cond Expansion.Conjunction Temporal.Asynchronous.Precedence Contingency.Cause.Result Temporal.Asynchronous.Precedence Temporal.Synchronous
thereafter	11	Temporal.Asynchronous.Precedence
thereby	9 3	Contingency.Cause.Result Expansion.Manner.Arg1-as-manner
therefore	26	Contingency.Cause.Result
though	91 128 16	Comparison.Concession.Arg1-as-denier Comparison.Concession.Arg2-as-denier Comparison.Contrast
thus	1 111	Contingency.Cause+Belief.Result+Belief Contingency.Cause.Result

till	1 4	Contingency.Negative-condition.Arg2-as-negCond Temporal.Asynchronous.Precedence
ultimately	1 1 1 15	Contingency.Cause.Reason Expansion.Conjunction Temporal.Asynchronous.Precedence Contingency.Cause.Result Temporal.Asynchronous.Precedence
unless	98	Contingency.Negative-condition.Arg2-as-negCond
until	17 143 2 7	Contingency.Condition.Arg2-as-cond Temporal.Asynchronous.Precedence Temporal.Asynchronous.Succession Contingency.Condition.Arg2-as-cond Temporal.Asynchronous.Succession
upon	1 2 2	Temporal.Asynchronous.Succession Contingency.Cause.Reason Temporal.Synchronous Contingency.Cause.Reason Temporal.Synchronous
whatever	4	Comparison.Concession.Arg1-as-denier
when	3 1 1 21 12 2 116 2 1 4 1 65 1 1 11 160 1 2 90 2 1 65 4 509	Comparison.Concession.Arg1-as-denier Comparison.Concession.Arg2-as-denier Comparison.Contrast Contingency.Cause.Reason Contingency.Condition+SpeechAct Contingency.Condition.Arg2-as-cond Expansion.Level-of-detail.Arg2-as-detail Contingency.Condition.Arg2-as-cond Expansion.Manner.Arg2-as-manner Temporal.Asynchronous.Precedence Contingency.Condition.Arg2-as-cond Temporal.Asynchronous.Precedence Temporal.Asynchronous.Succession Contingency.Cause+Belief.Reason+Belief Temporal.Asynchronous.Succession Contingency.Cause.Reason Temporal.Asynchronous.Succession Contingency.Cause.Result Temporal.Asynchronous.Succession Contingency.Condition+SpeechAct Temporal.Asynchronous.Succession Contingency.Condition.Arg2-as-cond Temporal.Asynchronous.Succession Temporal.Synchronous Comparison.Contrast Temporal.Synchronous Contingency.Cause+Belief.Reason+Belief Temporal.Synchronous Contingency.Cause.Reason Temporal.Synchronous Contingency.Cause.Result Temporal.Synchronous Contingency.Condition+SpeechAct Temporal.Synchronous Contingency.Condition.Arg2-as-cond Temporal.Synchronous Expansion.Level-of-detail.Arg2-as-detail Temporal.Synchronous
when and if	1	Temporal.Asynchronous.Succession Contingency.Condition.Arg2-as-cond
whenever	9	Contingency.Condition.Arg2-as-cond
where	2	Contingency.Condition.Arg2-as-cond
whereas	5	Comparison.Contrast
whether	7	Comparison.Concession.Arg1-as-denier

while	203	Comparison.Concession.Arg1-as-denier
	2	Comparison.Concession.Arg2-as-denier
	140	Comparison.Contrast
	1	Comparison.Similarity
	43	Expansion.Conjunction
	27	Temporal.Synchronous Comparison.Concession.Arg1-as-denier
	5	Temporal.Synchronous Comparison.Concession.Arg2-as-denier
	197	Temporal.Synchronous Comparison.Contrast
	22	Temporal.Synchronous Expansion.Conjunction
	163	Temporal.Synchronous
with	2	Comparison.Concession.Arg1-as-denier
	1	Comparison.Contrast
	5	Contingency.Cause+Belief.Reason+Belief Expansion.Level-of-detail.Arg2-as-detail
	5	Contingency.Cause+Belief.Reason+Belief
	3	Contingency.Cause.Reason Expansion.Level-of-detail.Arg2-as-detail
	109	Contingency.Cause.Reason
	2	Contingency.Condition.Arg2-as-cond
	41	Expansion.Conjunction
	7	Expansion.Instantiation.Arg2-as-instance
	111	Expansion.Level-of-detail.Arg2-as-detail
	6	Expansion.Manner.Arg2-as-manner
5	Temporal.Synchronous	
without	19	Comparison.Concession.Arg2-as-denier
	1	Contingency.Cause.Reason
	1	Contingency.Cause.Result
	9	Contingency.Negative-condition.Arg2-as-negCond
	2	Expansion.Level-of-detail.Arg2-as-detail
	62	Expansion.Manner.Arg2-as-manner
yet	96	Comparison.Concession.Arg2-as-denier
	4	Comparison.Contrast
	2	Expansion.Conjunction

## B Explicit Connectives and their Modifiers

This section shows the modified and variant forms for each explicit connective.

about	2	about
accordingly	5	accordingly
additionally	7	additionally
after	1	18 months after
	1	25 years after
	1	29 years and 11 months to the day after
	4	a day after
	1	a few days after
	1	a few hours after
	1	a few months after
	1	a few weeks after
	2	a month after
	1	a week after
	1	a year after
	1	a year and a half after
	1	about a week after
	1	about three weeks after
	501	after
	1	almost immediately after
	1	eight months after
	2	especially after
	8	even after
	1	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
	2	just after
	1	just five months after
	1	just minutes after
	1	less than a month after
	1	long after
	2	minutes after
	2	months after
	1	more than a year after
	1	nearly a year and a half after
	1	nearly two months after
	3	one day after
	9	only after
	1	only three years after
	1	only two weeks after
	1	particularly after
	1	reportedly after

after (cont.)	1 1 9 1 1 1 7 1 1 1 1 2 1 1 1	right after seven years after shortly after six years after some time after sometimes after soon after three months after three years after two days after two months after two weeks after within a year after within minutes after years after
afterward	5 1	afterward shortly afterward
afterwards	4 1	afterwards shortly afterwards
albeit	1	albeit
along with	2	along with
also	1737	also
alternatively	6	alternatively
although	328	although
and	6402	and
and then	1	and then
and/or	2	and/or
as	794 3 13 33 6 2	as especially as even as just as much as particularly as
as a result	77 1	as a result largely as a result
as an alternative	2	as an alternative
as if	16	as if
as long as	23 1	as long as only as long as
as much as	4	as much as
as soon as	19 1	as soon as just as soon as
as though	5	as though
as well	18	as well
as well as	7	as well as
at the same time	65	at the same time

because	1 1 1 763 1 1 11 8 12 5 1 3 3 2 15 2 1 2 4	'cause apparently because at least partly because because especially because in large part because in part because just because largely because mainly because merely because not because only because particularly because partly because perhaps because presumably because primarily because simply because
because of	8 1 3	because of not because of partly because of
before	1 1 1 1 1 1 1 1 280 1 14 1 1 1 1 5 1 1 1 3 1 4 1 1 1 1	a day or two before a decade before a few weeks before a full five minutes before a week before about six months before almost before an average of six months before before decades before even before even before then five minutes before fully eight months before in the 3 1/2 years before just before just days before just eight days before long before several months before shortly before two days before two months before two years before years before
before and after	1	before and after

besides	19	besides
beyond	2	beyond
both+and	6	both+and
but	3731	but
but also	1	but also
but then	3	but then
but then again	1	but then again
by	424	by
	1	just by
	1	merely by
	1	mostly by
	4	only by
	1	particularly by
	1	partly by
	1	probably by
	1	simply by
by comparison	11	by comparison
by contrast	28	by contrast
by then	7	by then
consequently	10	consequently
conversely	2	conversely
depending on	3	depending on
depending upon	1	depending upon
despite	9	despite
due to	1	due to
earlier	15	earlier
either+or	38	either+or
else	1	else
except	12	except
finally	32	finally
for	60	for
for example	200	for example
for instance	98	for instance
from	1	far from
	6	from
further	7	further
furthermore	12	furthermore

given	9	given
hence	5	hence
however	491	however
if	6 87 1144 3 1	especially if even if if particularly if typically, if
if and when	3	if and when
if only	6	if only
if+then	38	if+then
in	70	in
in addition	165	in addition
in any case	3	in any case
in case	5 1	in case just in case
in contrast	12	in contrast
in fact	85	in fact
in order	55	in order
in other words	17	in other words
in particular	15	in particular
in short	4	in short
in sum	2	in sum
in that	1	in that
in the end	11	in the end
in the meantime	14	in the meantime
in the meanwhile	1	in the meanwhile
indeed	103	indeed
insofar as	2	insofar as
instead	112	instead
instead of	43	instead of
later	1 1 1 88 1	a half-hour later a week later five years later later three days later
later on	2	later on
lest	2	lest
like	4 1	like not like
likewise	8	likewise
meantime	2	meantime
meanwhile	192	meanwhile
more accurately	1	more accurately
moreover	101 2	moreover moreover
much less	3	much less
neither+nor	3	neither+nor
nevertheless	44	nevertheless

next	5	Next
no matter	8	no matter
nonetheless	27	nonetheless
nor	33	nor
not just+but	1	not just but
not just+but+also	1	not just+but+also
not only	5	not only
not only+also	1	not only+also
not only because of	1	not only because of
not only+but	19	not only+but
not only+but also	10	not only+but also
not so much as	1	not so much as
now that	21	now that
on	1	on
on the contrary	4	on the contrary
on the one hand+on the other	1	on the one hand+on the other
on the one hand+on the other hand	2	on the one hand+on the other hand
on the other hand	36	on the other hand
once	84	once
only	7	only
only if	13	only if
or	356	or
or otherwise	2	or otherwise
otherwise	19	otherwise
plus	1	plus
previously	54	previously
rather	17	rather
rather than	40	rather than
regardless	2	regardless
regardless of	6	regardless of
separately	74	separately
similarly	18	similarly
simultaneously	1 5	almost simultaneously simultaneously
since	2 8 5 185	especially since ever since particularly since since
since before	1	since before
so	9 267	even so so
so as	3	so as
so long as	4	so long as
so much as	1	so much as
so that	31	so that
specifically	10	specifically
still	1 192	even still still
subsequently	3	subsequently
such as	2	such as

that is	4	that is
then	2 339	even then then
thereafter	4 7	shortly thereafter thereafter
thereby	12	thereby
therefore	26	therefore
though	95 235	even though though
thus	112	thus
till	1 1 3	'til til till
ultimately	18	ultimately
unless	98	unless
until	3 1 1 164	at least until just until only until until
upon	5	upon
whatever	4	whatever
when	1 2 1 4 14 6 9 2 1 1050	at least not when at least when back when especially when even when just when only when particularly when usually when when
when and if	1	when and if
whenever	9	whenever
where	2	where
whereas	5	whereas
whether	7	whether
while	3 803	even while while
with	1 2 1 295	especially with even with particularly with with
without	1 1 92	at least not without perhaps without without
yet	102	yet

## C Implicit Relation Senses and Implicit Connectives

This appendix shows the sense distribution of the total of 21782 Implicit relation tokens in PDTB-3, along with the set of implicit connectives inserted to express that relation. Counts are provided for the sense as well as for each connective inserted for that sense.

Sense	No.	Implicit connectives (count)
Comparison.Concession+SpeechAct.Arg2-as-denier+SpeechAct	10	although (1), but (5), or (4)
Comparison.Concession.Arg1-as-denier	56	although (33), but (2), despite being (1), even though (15), granted (1), however (3), while (1)
Comparison.Concession.Arg2-as-denier	1438	although (135), but (542), by comparison (2), by contrast (2), despite (2), despite this (3), even though (25), furthermore (1), granted (1), however (597), in fact (1), nevertheless (54), nonetheless (8), regardless (1), still (8), though (3), whereas (1), while (6), yet (46)
Comparison.Contrast	983	although (9), and (1), but (177), by comparison (187), by contrast (118), conversely (2), however (77), in comparison (5), in comparison to the fact (1), in contrast (197), in fact (2), meanwhile (4), nevertheless (1), on the contrary (7), on the other hand (30), regardless (1), still (1), when (1), whereas (77), while (82), yet (3)
Comparison.Similarity	31	also (1), in other words (1), likewise (1), similarly (28)
Contingency.Cause+Belief.Reason+Belief	135	after all (1), as (15), as a result of (3), as a result of being (3), as evidence (1), because (78), because of (1), because of being (1), considering that (1), given (3), given that (1), in (1), in fact (6), inasmuch as (4), indeed (6), since (5), so (3), This is because (1), thus (1)
Contingency.Cause+Belief.Result+Belief	67	as a result (6), as such (2), consequently (6), in fact (1), so (32), therefore (16), thus (4)
Contingency.Cause+SpeechAct.Reason+SpeechAct	13	because (13)
Contingency.Cause+SpeechAct.Result+SpeechAct	9	in short (1), so (8)
Contingency.Cause.Reason	2950	although (1), as (336), as a result of (100), as a result of being (84), as a result of having (1), because (2059), because it was (1), because of (66), because of being (4), for (2), for example (1), for one thing (1), for the reason that (1), given (46), given that (2), however (1), in (4), in fact (3), in other words (1), in short (2), inasmuch as (12), indeed (3), insofar as (2), it is because (3), on account of being (1), since (206), so (1), so as (1), specifically (1), that is (1), This is because (1), with (2)
Contingency.Cause.Result	2835	accordingly (86), and (2), as (1), as a consequence (2), as a result (761), as it turns out (1), as such (5), because of that (4), but (7), consequently (213), finally (1), for that reason (2), furthermore (1), hence (16), in fact (2), in other words (1), in response (2), in short (1), in the end (4), indeed (1), so (988), so as (7), so that (2), that is (3), then (3), therefore (326), thus (388), thus being (1), to this end (1), ultimately (3)

Sense	No.	Implicit connectives (No.)
Contingency.Condition+SpeechAct	2	so (1), were one to (1)
Contingency.Condition.Arg1-as-cond	3	then (3)
Contingency.Condition.Arg2-as-cond	196	but (1), if (12), if he is (8), if I am (1), if it is (39), if it was (2), if it were (2), if one is (11), if one is to (1), if one were (1), if s/he is (1), if she is (2), if they are (45), if they were (6), if we are (9), if we were (2), if you are (5), if you were (1), in order (1), so as (18), when (27), with (1)
Contingency.Purpose.Arg1-as-goal	3	by (1), for that purpose (1), through (1)
Contingency.Purpose.Arg2-as-goal	1370	for (7), for the purpose of (11), in (9), in order (1267), in order (for them) (1), in order to (1), in order to be (3), so as (11), so that (5), thus (1), with the goal (1), with the goal of (1), with the purpose of (52)
Expansion.Conjunction	4386	accordingly (1), additionally (13), after (1), also (456), although (3), and (1862), as (11), as a matter of fact (1), as it turns out (1), at the same time (1), besides (14), but (40), by comparison (2), first (11), for instance (1), for one thing (1), further (55), furthermore (356), however (13), in addition (248), in addition to (2), in addition to being (1), in fact (434), in other words (3), in particular (1), in response (1), in return (1), in short (6), in sum (6), in the end (20), incidentally (1), indeed (117), instead (1), likewise (18), meanwhile (85), moreover (1), moreover (144), nevertheless (1), on the whole (10), or (6), overall (11), plus (5), second (1), separately (3), similarly (56), so (3), specifically (6), that is (2), then (3), third (1), what's more (1), whereas (1), while (341), with (1), yet (1)
Expansion.Disjunction	30	or (30)
Expansion.Equivalence	336	in fact (5), in other words (230), in short (20), in sum (4), in the end (1), indeed (41), namely (2), or (1), rather (1), specifically (3), that is (28)
Expansion.Exception.Arg1-as-excpt	3	alternatively (1), otherwise (2)
Expansion.Exception.Arg2-as-excpt	2	however (1), rather (1)
Expansion.Instantiation.Arg1-as-instance	1	generally (1)
Expansion.Instantiation.Arg2-as-instance	1532	and (1), as part of that (2), first (1), for example (786), for instance (605), for one (1), for one thing (13), in fact (17), in particular (67), in this case (1), indeed (10), on the one hand (2), specifically (24), while (1), with (1)
Expansion.Level-of-detail.Arg1-as-detail	256	by being (1), in (47), in being (3), in fact (6), in general (3), in other words (25), in short (79), in sum (26), in summary (3), in the end (3), indeed (22), on the whole (19), overall (6), rather (1), so (3), specifically (1), that is (6), ultimately (2)

Sense	No.	Implicit connectives (No.)
Expansion.Level-of-detail.Arg2-as-detail	3105	accordingly (1), after all (1), although (1), and (6), as (9), as a result (1), because (2), but (3), by (1), eventually (1), first (13), for example (98), for instance (40), for one thing (8), further (1), furthermore (4), however (1), in (8), in addition (1), in fact (355), in more detail (1), in other words (16), in particular (657), in short (18), in sum (1), in the end (4), indeed (147), instead (1), more specifically (18), more to the point (1), moreover (3), namely (28), on the contrary (1), on the one hand (2), on the whole (8), or (1), overall (4), particularly (5), since (2), so (1), specifically (1565), that is (49), then (3), thus (2), ultimately (3), when (1), with (8)
Expansion.Manner.Arg1-as-manner	572	by doing so (1), in (1), thereby (570)
Expansion.Manner.Arg2-as-manner	167	and (1), by (156), by being (6), by means of (1), in that (1), specifically (1), with (1)
Expansion.Substitution.Arg2-as-subst	450	but (1), instead (302), rather (147)
Temporal.Asynchronous.Precedence	1081	afterward (1), afterwards (6), and (4), as of now (1), before (5), eventually (9), finally (11), in the end (5), later (13), next (8), now (7), previously (1), since then (7), soon (1), subsequently (56), then (933), thereafter (3), ultimately (8), when (1), while (1)
Temporal.Asynchronous.Succession	208	after (17), after being (4), after having (1), and (1), before (4), before that (1), earlier (28), first (1), previously (115), prior to this (1), upon (2), upon being (1), when (31), with (1)
Temporal.Synchronous	539	all the while (1), and (2), as (22), at that time (3), at the same time (33), at the time (22), during that time (1), in (5), in the meantime (1), meanwhile (117), simultaneously (5), when (30), while (288), while being (5), with (4)
<b>TOTAL</b>	<b>21781</b>	

## D AltLex Senses

Sense	Num
Comparison.Concession+SpeechAct.Arg2-as-denier+SpeechAct	1
Comparison.Concession.Arg1-as-denier	4
Comparison.Concession.Arg2-as-denier	10
Comparison.Contrast	39
Comparison.Similarity Expansion.Manner.Arg2-as-manner	1
Comparison.Similarity	11
Contingency.Cause+Belief.Reason+Belief	6
Contingency.Cause+Belief.Result+Belief	2
Contingency.Cause.Reason Expansion.Instantiation.Arg2-as-instance	8
Contingency.Cause.Reason Expansion.Level-of-detail.Arg2-as-detail	3
Contingency.Cause.Reason	270
Contingency.Cause.Result Expansion.Manner.Arg1-as-manner	1
Contingency.Cause.Result	625
Contingency.Condition.Arg1-as-cond	4
Contingency.Condition.Arg2-as-cond	26
Contingency.Negative-condition.Arg2-as-negCond	2
Contingency.Purpose.Arg1-as-goal	2
Contingency.Purpose.Arg2-as-goal Expansion.Manner.Arg1-as-manner	1
Contingency.Purpose.Arg2-as-goal	31
Expansion.Conjunction	138
Expansion.Equivalence	10
Expansion.Exception.Arg2-as-excpt	3
Expansion.Instantiation.Arg1-as-instance	1
Expansion.Instantiation.Arg2-as-instance	52
Expansion.Level-of-detail.Arg1-as-detail	16
Expansion.Level-of-detail.Arg2-as-detail	42
Expansion.Substitution.Arg1-as-subst	27
Expansion.Substitution.Arg2-as-subst	2
Temporal.Asynchronous.Precedence Comparison.Concession.Arg2-as-denier	16
Temporal.Asynchronous.Precedence Comparison.Contrast	1
Temporal.Asynchronous.Precedence	91
Temporal.Asynchronous.Succession Comparison.Contrast	1
Temporal.Asynchronous.Succession Expansion.Level-of-detail.Arg2-as-detail	1
Temporal.Asynchronous.Succession	21
Temporal.Synchronous	29
TOTAL	1498

## E AltLexC Senses

Sense	Num
Comparison.Concession.Arg1-as-denier	9
Comparison.Similarity	51
Contingency.Cause+Belief.Result+Belief	3
Contingency.Cause.negResult	4
Contingency.Cause.Result	28
Contingency.Condition.Arg2-as-cond	44
Contingency.Purpose.Arg2-as-goal	1
Expansion.Conjunction	1
TOTAL	140

## F Explicit Relations Linked together with Other Explicit

The senses of explicit connectives jointly linking the same spans, recorded in the corpus using the Link notation (Section 8.5).

Explicit sense(s)	Explicit sense(s)	count
COMP.CONC+SPEECHACT.ARG2-AS-DENIER+SA	EXP.SUBST.ARG2-AS-SUBST	2
COMP.CONC.ARG2-AS-DENIER	COMP.CONC.ARG2-AS-DENIER	17
	COMP.CONTRAST	3
	CONT.CAUSE.REASON	2
	CONT.CONDITION.ARG1-AS-COND	1
	CONT.CONDITION.ARG2-AS-COND	3
	EXP.EXCEPTION.ARG1-AS-EXCPT	1
	EXP.LEVEL-OF-DETAIL.ARG2-AS-DETAIL	2
	EXP.SUBST.ARG2-AS-SUBST	7
	TEMP.ASYNCH.PRECEDENCE	36
	TEMP.ASYNCH.SUCCESION	3
	TEMP.SYNCHRONOUS	7
COMP.CONC.ARG2-AS-DENIER	TEMP.SYNCHRONOUS/CONT.CONDITION.ARG2-AS-COND	1
COMP.CONTRAST	COMP.CONTRAST	1
	EXP.SUBST.ARG2-AS-SUBST	1
	TEMP.ASYNCH.PRECEDENCE	6
CONT.CAUSE.REASON / COMP.CONC.ARG2-AS-DENIER	TEMP.ASYNCH.SUCCESION/CONT.CAUSE.REASON	1
CONT.CAUSE.RESULT	EXP.CONJUNCTION	57
	EXP.SUBST.ARG2-AS-SUBST	1
	TEMP.ASYNCH.PRECEDENCE	1
CONT.CONDITION.ARG2-AS-COND	TEMP.ASYNCH.PRECEDENCE	1
CONT.PURPOSE.ARG2-AS-GOAL	TEMP.ASYNCH.PRECEDENCE	1
EXP.CONJUNCTION	COMP.CONC.ARG2-AS-DENIER	72
	COMP.CONTRAST	10
	CONT.CAUSE.REASON	2
	EXP.CONJUNCTION	46
	EXP.INSTANTIATION.ARG2-AS-INSTANCE	2
	EXP.LEVEL-OF-DETAIL.ARG2-AS-DETAIL	6
	EXP.SUBST.ARG2-AS-SUBST	9
	TEMP.ASYNCH.PRECEDENCE	123
	TEMP.ASYNCH.PRECEDENCE/COMP.CONC.ARG2-AS-DENIER	1
	TEMP.ASYNCH.SUCCESION	3
TEMP.SYNCHRONOUS	15	
EXP.SUBST.ARG2-AS-SUBST	EXP.DISJUNCTION	2
TEMP.ASYNCH.PRECEDENCE	TEMP.ASYNCH.PRECEDENCE	2
	TEMP.ASYNCH.PRECEDENCE/CONT.CAUSE.RESULT	1
TEMP.SYNCHRONOUS	TEMP.SYNCHRONOUS	2

## G Explicit Relations Linked with Implicits

This section provides the distribution of the explicit relations that are linked with implicit relations over the same argument spans.

Explicit Conn	Explicit Sense(s)	Implicit Sense(s)
also	Exp.Conjunction	Comp.Concession.Arg2-as-denier (3), Comp.Contrast (2), Temp.Synchronous (1)
and	Comp.Contrast	Exp.Substitution.Arg2-as-subst (60)
	Cont.Purpose.Arg2-as-goal	Exp.Manner.Arg1-as-manner (21)
	Exp.Conjunction	Comp.Concession.Arg2-as-denier (45), Comp.Contrast (39), Comp.Similarity (10), Cont.Cause.Result (760), Cont.Cause.Result Exp.Manner.Arg1-as-manner (47), Cont.Cause+Belief.Result+Belief (35), Cont.Purpose.Arg1-as-goal (1), Cont.Purpose.Arg2-as-goal Exp.Manner.Arg1-as-manner (2), Exp.Equivalence (8), Exp.Inst.Arg2-as-instance (15), Exp.Level-of-detail.Arg2-as-detail (60), Exp.Manner.Arg1-as-manner (21), Temp.Asynch.Precedence (432), Temp.Asynch.Precedence Comp.Concession.Arg2-as-denier (1), Temp.Asynch.Precedence Exp.Subst.Arg2-as-subst (1), Temp.Asynch.Succession (1), Temp.Synchronous (25)
as well	Comp.Similarity	Cont.Cause.Result Exp.Manner.Arg1-as-manner (1)
at the same time	Temp.Synchronous	Comp.Concession.Arg2-as-denier (5), Comp.Contrast (2)
but	Comp.Concession.Arg2-as-denier	Temp.Asynch.Precedence (13), Exp.Substitution.Arg2-as-subst (23), Exp.Exception.Arg1-as-excpt (1)
earlier	Temp.Asynch.Succession	Comp.Contrast (1)
even then	Temp.Asynch.Precedence  Comp.Concession.Arg2-as-denier	Comp.Contrast (1)
for example	Exp.Inst.Arg2-as-instance	Cont.Cause.Result (2)
in fact	Comp.Contrast	Exp.Substitution.Arg2-as-subst (4)
	Exp.Level-of-detail.Arg2-as-detail	Exp.Substitution.Arg2-as-subst (1)
in order	Cont.Purpose.Arg2-as-goal	Exp.Manner.Arg1-as-manner (6), Exp.Substitution.Arg2-as-subst (1)
indeed	Exp.Conjunction	Exp.Inst.Arg2-as-instance (1)
	Exp.Level-of-detail.Arg2-as-detail	Cont.Cause.Reason (1)
instead	Exp.Substitution.Arg2-as-subst	Comp.Contrast (2), Cont.Cause.Reason (1)
later	Temp.Asynch.Precedence	Comp.Concession.Arg2-as-denier (2), Cont.Cause.Result (2), Exp.Conjunction (1)
otherwise	Cont.Neg-cond.Arg1-as-negCond	Cont.Cause.Reason (2)
previously	Temp.Asynch.Succession	Comp.Contrast (1)
rather	Exp.Substitution.Arg2-as-subst	Comp.Contrast (2)
then	Cont.Cause.Result	Cont.Cause.Reason (2)
	Temp.Asynch.Precedence	Exp.Conjunction (19), Comp.Concession.Arg2-as-denier (7), Comp.Contrast (5), Cont.Cause.Reason (1)

Explicit Conn	Explicit Sense(s)	Implicit Sense(s)
thus	Cont.Cause.Result	Exp.Conjunction (1)
ultimately	Temp.Asynch.Precedence	Cont.Cause.Reason (1)
when	Temp.Synchronous	Exp.Level-of-detail.Arg2-as-detail (1)
with	Exp.Conjunction	Temp.Asynch.Precedence (5)
	Exp.Level-of-detail.Arg2-as-detail	Temp.Asynch.Precedence (1)
without	Exp.Manner.Arg2-as-manner	Comp.Concession.Arg2-as-denier (21)

## H AltLex Relations Linked with Explicits

This section provides the distribution of AltLex relations that are linked with explicit relations over the same argument spans. (Below, NegCond2 is an abbreviation of Arg2-as-negCond.)

AltLex Sense	AltLex	Explicit Sense	Explicit Conn
Comp.Conc.Arg2-as-denier	that's all (1)	Exp.Conjunction	and
Comp.Contrast	the swift passage contrasted with (1)	Exp.Conjunction	and
Comp.Similarity	equally (1), similar (1)	Exp.Conjunction	and
Cont.Cause.Reason	that's largely due to (1)	Comp.Conc.Arg2-as-denier	but
Cont.Cause.Reason	attributed that to (1), warned (1)	Exp.Conjunction	and
Cont.Cause+Belief.Result+Belief	in that context (1)	Exp.Conjunction	and
Cont.Neg-condition.negCond2	that's without (1)	Exp.Conjunction	and
Cont.Purpose.Arg1-as-goal	for that purpose (1)	Exp.Conjunction	and
Cont.Cause.Result	as such (1), because of the planned sale (1), for it (1), for that reason (2), it will cause (1), led (1), result in (1), resulted . . . in (1), resulted in (1), such prepayment forces (1), that includes (1), that is one reason why (1), that is why (1), that means (2), that permits (1), that resulted in (1), that will translate into (1), that would only strengthen (1), that's why (1), the battle has produced (1), therein (1), this has resulted in (1), will result in (1)	Exp.Conjunction	and
Cont.Cause.Result	blurring (1), bringing (1), damping (1), easing (1), forcing (1), giving (1), limiting (1), making (1), spreading (1)	Cont.Cause.Result	thus
Exp.Conjunction	further (1)	Exp.Conjunction	and
Exp.Level-of-detail.Arg2-as-detail	cites (1)	Exp.Conjunction	and
Exp.Instantiation.Arg1-as-instance	in general (1)	Comp.Conc.Arg2	but
Exp.Instantiation.Arg2-as-instance	in one instance (1)	Exp.Conjunction	and
Exp.Level-of-detail.Arg1-as-detail	in general (1)	Exp.Conjunction	and
Exp.Level-of-detail.Arg1-as-detail	in general (1)	Comp.Conc.Arg1-as-denier	though
Exp.Substitution.Arg1-as-subst	not (6), wasn't (1)	Exp.Conjunction	and
Temp.Asynch.Precedence	after that (1)	Exp.Conjunction	and
Temp.Asynch.Precedence	after that (1), since then (1)	Comp.Conc.Arg2-as-denier	but
Temp.Asynch.Precedence	followed by (1)	Temp.Asynch.Precedence	3 days later
Temp.Asynch.Precedence	it was followed by (1), now (4), since then (2), soon (1), this time around (1)	Exp.Conjunction	and
Temp.Asynch.Precedence	subsequent (1)	Exp.Conjunction	also
Temp.Asynch.Succession	before that (2)	Exp.Conjunction	and
Temp.Synchronous	at this point (1)	Exp.Conjunction	also
Temp.Synchronous	while she's doing it (1)	Exp.Conjunction	and

## I AltLex Relations Linked with Implicits

This section provides the distribution of relations with alternative lexicalizations that are linked with implicit relations over the same argument spans.

AltLex Sense	AltLex	Implicit Sense(s)
Contingency.Cause.Result	this can avoid	Comparison.Contrast (1)
Contingency.Purpose.Arg2-as-goal	in an apparent attempt in an attempt seeking	Expansion.Manner.Arg1-as-manner (1) Expansion.Manner.Arg1-as-manner (1) Expansion.Manner.Arg1-as-manner (1)
Expansion.Instantiation.Arg2-as-instance	by by means of	Expansion.Manner.Arg2-as-manner (1) Expansion.Manner.Arg2-as-manner (1)
Expansion.Substitution.Arg1-as-subst	not	Comparison.Contrast (1)
Temporal.Asynch.Precedence	now soon that followed that was before now next	Comparison.Conc.Arg2-as-denier (2) Comparison.Conce.Arg2-as-denier (1) Comparison.Conc.Arg2-as-denier (1) Comparison.Conc.Arg2-as-denier (1) Comparison.Contrast (2) Contingency.Cause.Result (1)
Temporal.Synchronous	in the same period	Comparison.Contrast (1)