

# Discourse Relations

*Adapted from the “Penn Discourse Treebank 2.0 Annotation Manual” by the PDTB Research Group, 2007*

What is a Discourse Relation? .....	2
What to Annotate: The Parts of a Discourse Relation.....	3
Relation Type.....	3
Arguments.....	3
Explicit Connective: Bare and Full .....	3
Implicit Connectives .....	3
Connective Senses .....	4
Supplements.....	4
Attribution.....	4
Attribution: Type .....	4
Attribution: Source.....	4
Attribution: Polarity .....	4
Attribution: Determinacy .....	4
Explicit Relations.....	5
Connectives.....	5
Location of Connectives .....	6
Bare vs. Full Connectives.....	7
Discontinuous Connectives .....	7
Arguments.....	7
Simple Clause Arguments.....	7
Multi-Clause or Sentential Arguments.....	8
Non-Clausal Arguments.....	8
Clause-internal Complements and Non-clausal Adjuncts .....	9
Implicit Relations.....	11
Arguments.....	12
Sub-sentential arguments .....	12
Multiple sentence arguments.....	12
Arguments involving parentheticals.....	13
Entity-Based Relations .....	14
Alternative Lexicalization Relations.....	15
“No Relation” Relations .....	16
Senses .....	17
Attribution .....	18
Differences between this guide and PDTB-2.0.....	19

## What is a Discourse Relation?

Most documents have some sort of discourse structure: a logical flow of events, states, and propositions that makes for a coherent idea, argument, or story. The Discourse Relation annotation scheme is designed to capture this structure by marking the so-called *discourse connectives* and their arguments, as well as some additional information regarding their meaning and attribution.

The following examples each fall into one of the four major types of discourse relationship we will mark:

- (1) I refused to pay the cobbler the full \$95 because he did poor work. (Contingency)
- (2) He knows a tasty meal when he eats one. (Temporal)
- (3) IBM's stock price rose, but the overall market fell. (Comparison)
- (4) I never gamble too far; in particular, I quit after one try. (Expansion)

In example (1) the narrator (“I”) is telling us the reason that he did not pay the cobbler: *because* the cobbler did poor work. The discourse connective “because” explicitly marks a discourse relation between two propositions, the first argument stating the refusal to pay, and the second argument stating the poor work. In later exam

**Notation:** We will underline discourse connectives, and give the sense of the connective at the end in parentheses. Arguments (and, later, supplements) will be surrounded by square brackets with an identifying subscript after the opening bracket. Arg1's will be additionally highlighted with italics, and Arg2's with bold face.

- (5) [<sub>ARG1</sub> *I refused to pay the cobbler the full \$95*] because [<sub>ARG2</sub> **he did poor work**]. (Contingency)

Discourse relations need not be explicit in the text; they may be *implicit*, as in this example:

- (6) They were getting way too much unwanted junk mail in their inbox. Now, all of those are automatically flagged as spam.

When indicating these types of examples in this guide, we will mark the arguments as we do for explicit relations, but will insert, in curly brackets, an inferred discourse connective, prefixed with “implicit=”.

- (7) [<sub>ARG1</sub> *They were getting way too much unwanted junk mail in their inbox.*] {Implicit=“so”} [<sub>ARG2</sub> **Now, all of those are automatically flagged as spam.**] (Contingency)

Here we have two arguments connective by an implicit discourse connective, which the annotator has identified as “so”.

There are many different ways of express types of discourse relationships. In particular, we will distinguish four different ways of expressing a discourse relationship: Explicit, Implicit, Alternative Lexicalization, and Entity-Based Coherence. We will also mark the absence of a discourse relationship (“No Relation” relations) between two arguments where a relation might otherwise be expected. As shown in examples (1) - (4), for Explicit and Implicit relationships, there are four major types, or *senses*: Temporal, Contingency, Comparison, and Expansion. Each of these senses is further split into types and subtypes, as discussed in later sections.

## What to Annotate: The Parts of a Discourse Relation

Every discourse relation has a type and two arguments. Depending on the relation type and how it is expressed, up to five additional groups of information might need to be annotated: the explicit or implicit connective; the connective senses; supplements to the arguments; and attribution information for the relation and the two arguments.

### Relation Type

There are four basic types of relation: Explicit, Implicit, Alternative Lexicalization, and Entity-Based Coherence. Additionally, we will be marking “No Relation” between two arguments where a relation would otherwise be expected. While every relation type takes two arguments, they differ in what other information is required. This is laid out in the following table.

Type	Explicit Connective	Implicit Connectives	Connective Senses	Argument Supplements	Attribution	Arguments
<b>Explicit</b>	Yes	-	Yes	Yes	Yes	Yes
<b>Implicit</b>	-	Yes	Yes	Yes	Yes	Yes
<b>AltLex</b>	Yes	-	Yes	Yes	Yes	Yes
<b>EntRel</b>	-	-	-	-	-	Yes
<b>NoRel</b>	-	-	-	-	-	Yes

### Arguments

Every discourse relation has two arguments. Arguments are abstract objects (sometimes abbreviated AO) such as events, states, and propositions. The two arguments to a discourse connective are simply labeled Arg2, for the argument that appears in the clause that is syntactically bound to the connective, and Arg1, for the other argument.

### Explicit Connective: Bare and Full

Explicit relations and Alternative Lexicalization relations require that the discourse connective be identified in the text: thus the connective is *explicit*. Connectives for Explicit relations are drawn primarily from well-defined syntactic classes: subordinating conjunctions (e.g., because, when, since, although), coordinating conjunctions (e.g., and, or, nor), adverbials (e.g., however, otherwise, then, as a result, for example). The simple form of the explicit connective is called the *bare* connective. Sometimes, an explicit relation may be modified, for example “*partly* because” or “*especially* when”. In these cases, the full connective includes the modifier, whereas the *bare* connective contains only the actual discourse connective.

In the case of Alternative Lexicalization relations, the annotator must mark a text span in Arg2 expressing how the relation is selected and marked. This is the equivalent of a bare explicit connective for the Alternatively Lexicalized relation.

### Implicit Connectives

Implicit relations have no text anchor for their connective, and therefore the annotator must supply at least one discourse connective (possibly two) that carries the same meaning that was inferred. Implicit connectives are used only for Implicit relations.

## Connective Senses

Sense labels, which are required for Explicit, Implicit, and Alternative Lexicalization relations, are drawn from a hierarchical classification - a three-level hierarchy grouping connectives into classes, types and subtypes.

## Supplements

Supplements to Arg1 and Arg2, called Sup1 for material supplementary to Arg1, and Sup2, for material supplementary to Arg2, are annotated to mark material that is relevant but not “minimally necessary” for interpreting the relation.

## Attribution

Attribution is the encoding of “ownership” between an individual and a relation or argument. Attribution contains four features: type, source, polarity, and determinacy. Attribution has to do with ascribing beliefs and assertions expressed in text to the agent(s) holding or making them.

Each attribution feature not only has a feature value (as indicated in the description below), but may also a textual span, which is a set of tokens in the text that actually express the value of the feature being captured. Attribution is only annotated on Explicit, Implicit, and Alternative Lexicalization relations and their arguments.

### Attribution: Type

The attribution type feature signifies the nature of the relation between an agent and a relation or argument, leading to different inferences about the degree of factuality of the AO. There are four distinct sub-types of attribution: Assertion propositions, Belief propositions, Facts and Eventualities. A relation or argument may also have a Null attribution type.

### Attribution: Source

The source feature tells us who is asserting the relation or argument, and distinguishes between four cases: (a) Writer: the writer of the text; (b) Other: some specific agent, other than the writer, introduced in the text; (c) Arbitrary: some arbitrary individual(s) indicated via a non-specific reference in the text; or (d) Inherited: an argument inherits its source from its parent relation.

### Attribution: Polarity

The scopal polarity feature is annotated on relations and arguments to identify cases where verbs of attribution are negated on the surface - syntactically (e.g., didn’t say, don’t think) or lexically (e.g., denied), but where the negation in fact reverses the polarity of the attributed relation or argument content. Scopal polarity may be either *true* or *false*.

### Attribution: Determinacy

The determinacy feature captures the fact that the attribution for a relation or argument can itself be cancelled in particular circumstances, such as within negated, conditional, and infinitive contexts. Determinacy may be either *true* or *false*.

## Explicit Relations

Explicit relations are discourse relations that have an explicit lexical marker (an actual word), drawn from specific grammatical categories, that indicates the relation in the text. Not all relations that have lexical markers are explicit – sometimes they are Alternative Lexicalizations (see the relevant section below).

## Connectives

Explicit connectives are drawn from the following grammatical classes:

- Subordinating conjunctions (e.g., *because, when, since, although*, as in (8) and (9))
- Coordinating conjunctions (e.g., *and, or, nor*, as in (10) and (11)):
- Adverbial and prepositional phrases (e.g., *however, otherwise, as a result*, as in (12) and (13)).

- (8) Since [ARG2 **McDonald’s menu prices rose this year**], [ARG1 *the actual decline may have been more*].
- (9) [ARG1 *The federal government suspended sales of U.S. savings bonds*] because [ARG2 **Congress hasn’t lifted the ceiling on government debt**].
- (10) [ARG1 *The House has voted to raise the ceiling to \$3.1 trillion*], but [ARG2 **the Senate isn’t expected to act until next week at the earliest**].
- (11) [ARG1 *Only 19% of the purchasing managers reported better export orders in October, down from 27% in September.*] And [ARG2 **8% said export orders were down last month, compared with 6% the month before**].
- (12) [ARG1 *Working Woman, with circulation near one million, and Working Mother, with 625,000 circulation, are legitimate magazine success stories.*] [ARG2 **The magazine Success**], however, [ARG2 **was for years lackluster and unfocused**].
- (13) [ARG1 *In the past, the socialist policies of the government strictly limited the size of new steel mills, petrochemical plants, car factories and other industrial concerns to conserve resources and restrict the profits businessmen could make.*] As a result, [ARG2 **industry operated out of small, expensive, highly inefficient industrial units**].

Adverbials that do not denote relations between two arguments have not been annotated as discourse connectives. For example, adverbials called “cue phrases” or “discourse markers” such as *well, anyway, now*, etc. should not be annotated since they serve to signal the organizational or focus structure of the discourse, rather than relate arguments. And clausal adverbials such as *strangely, probably, frankly, in all likelihood*, etc. should also not be annotated as discourse connectives since they take a single argument, rather than two.

Not all tokens of words and phrases that can serve as Explicit connectives actually do so: Some tokens serve other functions, such as to relate non-discourse-argument entities (e.g., the use of *and* to conjoin noun phrases in (14), and the use of *for example* to modify a noun phrase in (15)), to relativize extracted adjuncts (e.g., the use of *when* to relativize the time NP in (16)), and so on (17). Such expressions are not annotated as discourse connectives.

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

- (15) These mainly involved such areas as materials—advanced soldering machines, *for example*—and medical developments derived from experimentation in space, such as artificial blood vessels.
- (16) Equitable of Iowa Cos., Des Moines, had been seeking a buyer for the 36-store Younkers chain since June, *when* it announced its intention to free up capital to expand its insurance business.
- (17) The products already available are cross-connect systems, used *instead of* mazes of wiring to interconnect other telecommunications equipment.

Conjoined connectives like *when and if* and *if and when* are annotated as a single connective. Examples are shown in (18) and (19).

- (18) When and if [ARG2 **the trust runs out of cash**] – which seems increasingly likely – [ARG1 *it will need to convert its Manville stock to cash*].
- (19) Hoylake dropped its initial \$13.35 billion (\$20.71 billion) takeover bid after it received the extension, but said [ARG1 *it would launch a new bid*] *if and when* [ARG2 **the proposed sale of Farmers to Axa receives regulatory approval**].

### Location of Connectives

Connectives and their arguments can appear in any relative order. For the subordinating conjunctions, since the subordinate clause is bound to the connective, Arg2 corresponds to the subordinate clause, and hence the linear order of the arguments can be Arg1-Arg2 (20), Arg2-Arg1 (21), or Arg2 may appear between discontinuous parts of Arg1 (22), depending on the relative position of the subordinate clause with respect to its matrix clause.

- (20) [ARG1 *The federal government suspended sales of U.S. savings bonds*] because [ARG2 **Congress hasn't lifted the ceiling on government debt**].
- (21) Because [ARG2 **it operates on a fiscal year**], [ARG1 *Bear Stearns's yearly filings are available much earlier than those of other firms*].
- (22) [ARG1 *Most oil companies*], when [ARG2 **they set exploration and production budgets for this year**], [ARG1 *forecast revenue of \$15 for each barrel of crude produced*].

The order of the arguments for adverbials and coordinating conjunctions is typically Arg1-Arg2 since Arg1 usually appears in the prior discourse. But as (23) shows, Arg1 of a discourse adverbial can also appear within Arg2, which is then annotated as two discontinuous spans.

- (23) As an indicator of the tight grain supply situation in the U.S., market analysts said that [ARG2 **late Tuesday the Chinese government**], [ARG1 *which often buys U.S. grains in quantity*], [ARG2 **turned**] instead [ARG2 **to Britain to buy 500,000 metric tons of wheat**].

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

- (24) [ARG1 *Despite the economic slowdown, there are few clear signs that growth is coming to a halt.*] As a result, [ARG2 **Fed officials may be divided over whether to ease credit**].
- (25) [ARG1 *The chief culprits, he says, are big companies and business groups that buy huge amounts of land “not for their corporate use, but for resale at huge profit.”*] [ARG2 **The Ministry of Finance**], as a result, [ARG2 **has proposed a series of measures that would restrict business investment in real estate.**]

- (26) [ARG1 *Polyvinyl chloride capacity “has overtaken demand”*] [ARG2 **and we are experiencing reduced profit margins**] as a result.”

### Bare vs. Full Connectives

Many connectives can occur with adverbs such as *only*, *even*, *at least*, and so on. We refer to such tokens as modified connectives (with the connective as head and the adverb as modifier). Some examples are given below, with the adverb shown in parentheses for clarity. Rather than distinguishing such occurrences as a separate type, they are treated as the same type as that of the head - the bare form. Appendix C in the full annotation guide gives a more comprehensive list of possible modified connectives for each connective type.

- (27) That power can sometimes be abused, (particularly) since jurists in smaller jurisdictions operate without many of the restraints that serve as corrective measures in urban areas.
- (28) You can do all this (even) if you’re not a reporter or a researcher or a scholar or a member of Congress.
- (29) We’re seeing it (partly) because older vintages are growing more scarce.

Modified connectives should be annotated as normal connectives, where the core connective is annotated as the *Bare* connective, and the whole connective, including modifier, is annotated as the *Full* connective.

### Discontinuous Connectives

In addition to modified forms of connectives, one might also have “parallel” connectives, that is, pairs of connectives where one part presupposes the presence of the other, and where both together take the same two arguments. Such instances should be annotated as a single, discontinuous connective.

- (30) On the one hand, Mr. Front says, [ARG1 *it would be misguided to sell into “a classic panic.”*] On the other hand, [ARG2 **it’s not necessarily a good time to jump in and buy**].
- (31) If [ARG1 *the answers to these questions are affirmative*], then [ARG2 **institutional investors are likely to be favorably disposed toward a specific poison pill**].
- (32) Either [ARG1 *sign new long-term commitments to buy future episodes*] or [ARG2 **risk losing “Cosby” to a competitor**].

### Arguments

The delineation of arguments is governed by *the Minimality Principle*: only as many clauses and/or sentences should be included in an argument selection as are minimally required and sufficient for the interpretation of the relation. Any other span of text that is perceived to be relevant (but not necessary) in some way to the interpretation of arguments is annotated as supplementary information, labeled Sup1 and Sup2, for Arg1 and Arg2 respectively.

There is no restriction on how far an argument can be from its corresponding connective. Arguments can be found in the same sentence as the connective, in the sentence immediately preceding that of the connective, or in some non-adjacent sentence.

### Simple Clause Arguments

With a few exceptions to be discussed below, the simplest syntactic realization of a connective’s argument is taken to be a clause, tensed or non-tensed. Further, the clause can be a matrix clause, a

complement clause, or a subordinate clause. Some examples of single clausal realizations are shown in the following examples.

- (33) A Chemical spokeswoman said [ARG1 *the second-quarter charge was*] [ARG2 **“not material” and that no personnel changes were made**] as a result.
- (34) In Washington, House aides said Mr. Phelan told congressmen that the collar, [ARG1 *which banned program trades through the Big Board’s computer*] when [ARG2 **the Dow Jones Industrial Average moved 50 points**], didn’t work well.
- (35) [ARG1 *Knowing a tasty—and free—meal*] when [ARG2 **they eat one**], the executives gave the chefs a standing ovation.
- (36) Alan Smith, president of Marks & Spencer North America and Far East, says that Brooks Brothers’ focus is to boost sales [ARG1 *by broadening its merchandise assortment*] while [ARG2 **keeping its “traditional emphasis.”**]
- (37) Radio Shack says it has a policy [ARG1 *against selling products*] if [ARG2 **a salesperson suspects they will be used illegally.**]
- (38) “We have been a great market [ARG1 *for inventing risks*] [ARG2 **which other people**] then [ARG2 **take, copy and cut rates**].”

### Multi-Clause or Sentential Arguments

In addition to single clauses, abstract object arguments of connectives can also be realized as multiple clauses and multiple sentences. Example (39) shows multiple sentences selected for the Arg1 argument of *still*. Multiple clause and multiple sentence arguments can also be annotated discontinuously if they so appear in the text.

- (39) [ARG1 *Here in this new center for Japanese assembly plants just across the border from San Diego, turnover is dizzying, infrastructure shoddy, bureaucracy intense. Even after-hours drag; “karaoke” bars, where Japanese revelers sing over recorded music, are prohibited by Mexico’s powerful musicians union.*] Still, [ARG2 **20 Japanese companies, including giants such as Sanyo Industries Corp., Matsushita Electronics Components Corp. and Sony Corp. have set up shop in the state of Northern Baja California.**]

There are no restrictions on how many or what types of clauses can be included in these complex selections, except for the Minimality Principle.

### Non-Clausal Arguments

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

*Verb Phrase coordinations:* While the conjunction in a coordinated verb phrase is not annotated as a distinct discourse connective, one or more verb phrases within the coordinated structure can be annotated as the argument of another connective. However, the subject of the verb phrase coordinates is included in the argument selection only for the first verb phrase coordinate (Arg1 of *then* in (40)). Subjects for non-initial coordinates are not included in the selection (Arg2 of *then* in (40) and Arg1 of *because* in (41)), and will have to be retrieved via independent heuristics to arrive at the complete interpretation of the argument.

- (40) [ARG1 *It acquired Thomas Edison’s microphone patent*] [ARG2 and] then [ARG2 **immediately sued the Bell Co.**] [SUP2 *claiming that the microphone invented by my grandfather, Emile*]

Berliner, which had been sold to Bell for a princely \$50,000, infringed upon Western Union's Edison patent].

- (41) She became an abortionist accidentally, [ARG1 *and continued*] because [ARG2 **it enabled her to buy jam, cocoa and other war-rationed goodies**].

*Nominalizations:* Nominations are annotated as arguments of connectives in two strictly restricted contexts. The first context is when they allow for an existential interpretation, as in (42), where the Arg1 selection can be interpreted existentially as that there will be major new liberalizations:

- (42) Economic analysts call his trail-blazing liberalization of the Indian economy incomplete, and many are hoping [ARG1 *for major new liberalizations*] if [ARG2 **he is returned firmly to power**].

The second context is when they involve a clearly observable case of a derived nominalization, as in (43), where the Arg1 selection can be assumed to be transformationally derived from such laws to be resurrected:

- (43) But in 1976, the court permitted [ARG1 *resurrection of such laws*], if [ARG2 **they meet certain procedural requirements**].

*Anaphoric expressions:* An anaphoric expression like *this* or *that* or *so* that refers to an argument can be annotated as Arg1 of a connective.

- (44) "It's important to share the risk [ARG1 *and even more so*] when [ARG2 **the market has already peaked**]."
- (45) Investors who bought stock with borrowed money – that is, "on margin" – may be more worried than most following Friday's market drop. [ARG1 *That's*] because [ARG2 **their brokers can require them to sell some shares or put up more cash to enhance the collateral backing their loans**].
- (46) Evaluations suggest that good ones are – [ARG1 *especially so*] if [ARG2 **the effects on participants are counted**].

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

- (47) Underclass youth are a special concern. [SUP1 *Are such expenditures worthwhile, then?*] [ARG1 *Yes*], if [ARG2 **targeted**].
- (48) [SUP1 *Is he a victim of Gramm-Rudman cuts?*] [ARG1 *No*], but [ARG2 **he's endangered all the same**]: His new sitcom on ABC needs a following to stay on the air.

### Clause-internal Complements and Non-clausal Adjuncts

All clauses that are selected as arguments of connectives, all complements of the main clausal predicate and all non-clausal adjuncts (e.g., a speciality chemicals concern in Arg2 of (49)), adverbs (e.g., for example in Arg1 of (50)), complementizers (e.g., that in Arg1 and Arg2 of (51)), conjunctions (e.g., But in Arg1 of (52)), and relative pronouns (e.g., whom in Arg1 of (53)) modifying the clause are obligatorily included in the argument (except for the connective that is itself being annotated), even if these elements are not necessary for the minimal interpretation of the relation.

- (49) Although [ARG2 **Georgia Gulf hasn't been eager to negotiate with Mr. Simmons and NL, a specialty chemicals concern**], [ARG1 *the group apparently believes the company's management is interested in some kind of transaction*].
- (50) Players must abide by strict rules of conduct even in their personal lives – [ARG1 *players for the Tokyo Giants, for example, must always wear ties*] when [ARG2 **on the road**].
- (51) There seems to be a presumption in some sectors of (Mexico's) government [ARG1 *that there is a lot of Japanese money waiting behind the gate*], and [ARG2 **that by slightly opening the gate, that money will enter Mexico**].
- (52) [ARG1 *But the Reagan administration thought otherwise*], and [ARG2 **so may the Bush administration**].
- (53) That impressed Robert B. Pamplin, Georgia-Pacific's chief executive at the time, [ARG1 *whom Mr. Hahn had met*] while [ARG2 **fundraising for the institute**].

Inclusion of non-clausal elements is obligatory even when it warrants discontinuous annotation:

- (54) They found students in an advanced class a year earlier who said she gave them similar help, [ARG1 *although*] because [ARG2 **the case wasn't tried in court**], [ARG1 *this evidence was never presented publicly*].
- (55) He says [ARG1 *that*] when [ARG2 **Dan Dorfman, a financial columnist with USA Today, hasn't returned his phone calls**], [ARG1 *he leaves messages with Mr. Dorfman's office saying that he has an important story on Donald Trump, Meshulam Riklis or Marvin Davis*].
- (56) Under two new features, participants will be able to transfer money from the new funds to other investment funds [ARG1 *or*], if [ARG2 **their jobs are terminated**], [ARG1 *receive cash from the funds*].
- (57) [ARG1 *Last week*], when [ARG2 **her appeal was argued before the Missouri Court of Appeals**], [ARG1 *her lawyer also relied on the preamble*].

## Implicit Relations

The goal of annotating Implicit relations in the is to capture relations between arguments that are not realized explicitly in the text and are left to be inferred by the reader. In (58), a causal relation is inferred between raising cash positions to record levels and high cash positions helping to buffer a fund, even though no explicit connective appears in the text to express this relation. Similarly, in (59), a consequence relation is inferred between the increase in the number of rooms and the increase in the number of jobs, though no explicit connective expresses this relation.

- (58) Several leveraged funds don't want to cut the amount they borrow because it would slash the income they pay shareholders, fund officials said. But a few funds have taken other defensive steps. [ARG1 *Some have raised their cash positions to record levels.*] {Implicit="because"} [ARG2 **High cash positions help buffer a fund when the market falls.**]
- (59) [ARG1 *The projects already under construction will increase Las Vegas's supply of hotel rooms by 11,795, or nearly 20%, to 75,500.*] {Implicit="so"} [ARG2 **By a rule of thumb of 1.5 new jobs for each new hotel room, Clark County will have nearly 18,000 new jobs.**]

Such inferred relations are annotated between adjacent sentences (regardless of whether they occur within the same paragraph) and are marked as Implicit connectives by specifying a connective expression that best expresses the inferred relation. So in the previous two examples, the Implicit connectives *because* and *so* are specified to capture the perceived causal and consequence relations, respectively. Multiple discourse relations between adjacent sentences may also be inferred, and should be annotated as multiple Implicit connectives. In (60), two Implicit connectives, *when* and *for example*, are inserted to express how Arg2 presents one instance of the circumstances under which Mr. Morishita comes across as an outspoken man of the world. Similarly, in (61), the two Implicit connectives *because* and *for example* are provided to express how Arg2 presents one instance of the reasons for the claim that the third principal did have garden experience.

- (60) [ARG1 *The small, wiry Mr. Morishita comes across as an outspoken man of the world.*] {Implicit="when", "for example"} [ARG2 **Stretching his arms in his silky white shirt and squeaking his black shoes he lectures a visitor about the way to sell American real estate and boasts about his friendship with Margaret Thatcher's son.**]
- (61) [ARG1 *The third principal in the S. Gardens adventure did have garden experience.*] {Implicit="because", "for example"} [ARG2 **The firm of Bruce Kelly/David Varnell Landscape Architects had created Central Park's Strawberry Fields and Shakespeare Garden.**]

The annotation of inferred relations should be done intuitively, by the annotator reading adjacent sentences (and in some cases, the preceding text as), and making a decision about whether or not a relation could be inferred between them, and providing an appropriate Implicit connective to express the inferred relation, if any. In cases where an Implicit connective could not be provided, there are three alternative markings: Alternative Lexicalization, Entity Relation, and "No Relation." See the relevant sections for descriptions

Implicit connectives should be annotated between all successive pairs of sentences (irrespective of paragraph boundaries), and should also be annotated intra-sententially if present.

## Arguments

Choosing arguments for implicit relations follow the same rules as those for explicit relations; this section covers a few additional points of relevance to arguments of implicit relations.

### Sub-sentential arguments

While implicit relations should be annotated between adjacent sentences when possible, this does not mean that all arguments of an implicit relation need span complete sentences. As with the explicit relations, you should select only as much text as is minimally necessary for the interpretation of the inferred relation. Furthermore, as for explicit relations, parts of the text that are relevant (but not necessary) to the interpretation of the relation should be marked as supplementary information. For instance, in (62), for the inferred exemplification relation, the matrix clause is excluded from Arg1, and is marked as Sup1 - its relevance being due to its containment of the referent of the relative pronoun when in Arg1.

- (62) [<sub>Sup1</sub> *Average maturity was as short as 29 days at the start of this year*], [<sub>ARG1</sub> *when short-term interest rates were moving steadily upward.*] {Implicit="for example"} [<sub>ARG2</sub> *The average seven- day compound yield of the funds reached 9.62% in late April.*]

Parts of the sentence may also be left out without being labeled as supplementary information, when it is not relevant to the interpretation of the relation, as for example, the non-restrictive relative clause in the sentence containing Arg2 in (63).

- (63) [<sub>ARG1</sub> *Meanwhile, the average yield on taxable funds dropped nearly a tenth of a percentage point, the largest drop since midsummer.*] {Implicit="in particular"} [<sub>ARG2</sub> **The average seven-day compound yield**], which assumes that dividends are reinvested and that current rates continue for a year, [<sub>ARG2</sub> **fell to 8.47%, its lowest since late last year, from 8.55% the week before, according to Donoghue's**].

Attribution (covered later) is also a cause for selection of sub-sentential spans, as seen in the sentence containing Arg1 in (64), and both the sentences containing Arg1 and Arg2 in (65).

- (64) “[<sub>ARG1</sub> *Lower yields are just reflecting lower short-term interest rates*],” said Brenda Malizia Negus, editor of Money Fund Report. {Implicit="since"} [<sub>ARG2</sub> **Money funds invest in such things as short-term Treasury securities, commercial paper and certificates of deposit, allof which have been posting lower interest rates since last spring.**]
- (65) Ms. Terry did [<sub>ARG1</sub> *say the fund's recent performance “illustrates what happens in a leveraged product” when the market doesn't cooperate*]. {Implicit="still"} “[<sub>ARG2</sub> **When the market turns around**],” she says, “[<sub>ARG2</sub> **it will give a nice picture” of how leverage can help performance**].

### Multiple sentence arguments

In addition to selecting sub-sentential clauses, either argument can also span over multiple sentences (discontinuously, if necessary) if such an extension is minimally required for the interpretation of the relation. For instance, for the inferred exemplification relation in (66), the example of legal controversies always assuming a symbolic significance far beyond the particular case is given not just by the sentence following it, but rather by a combination of the three following sentences.

- (66) Legal controversies in America have a way of assuming a symbolic significance far exceeding what is involved in the particular case. They speak volumes about the state of our society at a

given moment. [ARG1 *It has always been so.*] {Implicit=“for example”} [ARG2 **In the 1920s, a young schoolteacher, John T. Scopes, volunteered to be a guinea pig in a test case sponsored by the American Civil Liberties Union to challenge a ban on the teaching of evolution imposed by the Tennessee Legislature. The result was a world-famous trial exposing profound cultural conflicts in American life between the “smart set,” whose spokesman was H.L. Mencken, and the religious fundamentalists, whom Mencken derided as benighted primitives. Few now recall the actual outcome: Scopes was convicted and fined \$100, and his conviction was reversed on appeal because the fine was excessive under Tennessee law.**]

Lists, when they span multiple sentences, are also taken to be minimal. Arg1 is extended to include the complete list:

- (67) All the while, Ms. Bartlett had been busy at her assignment, serene in her sense of self-tith. [ARG1 *As she put it in a 1987 lecture at the Harvard Graduate School of Design: “I have designed a garden, not knowing the difference between a rhododendron and a tulip.” Moreover, she proclaimed that “landscape architects have been going wrong for the last 20 years” in the design of open space. And she further stunned her listeners by revealing her secret garden design method: Commissioning a friend to spend “five or six thousand dollars . . . on books that I ultimately cut up.” After that, the layout had been easy. “I’ve always relied heavily on the grid and found it never to fail.”*] {Implicit=“in addition”} [ARG2 **Ms. Bartlett told her audience that she absolutely did not believe in compromise or in giving in to the client “because I don’t think you can do watered-down versions of things.”**]

The next shows multiple sentences selected for both Arg1 and Arg2, as minimally required for Arg2, and as a list for Arg1.

- (68) While the model was still on view, [ARG1 *Manhattan Community Board 1 passed a resolution against South Gardens. The Parks Council wrote the BPCA that this “too ‘private’ . . . exclusive,” complex and expensive “enclosed garden . . . belongs in almost any location but the water-front.”*] {Implicit=“similarly”} [ARG2 **Lynden B. Miller, the noted public garden designer who restored Central Park’s Conservatory Garden, recalls her reaction to the South Gardens model in light of the public garden she was designing for 42<sup>nd</sup> Street’s Bryant Park: “Bryant Park, as designed in 1933, failed as a public space, because it made people feel trapped. By removing the hedges and some walls, the Bryant Park Restoration is opening it up. It seems to me the BPCA plan has the potential of making South Gardens a horticultural jail for people and plants.”**]

### Arguments involving parentheticals

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

## Entity-Based Relations

Entity-based relations capture cases where the implicit relation between adjacent sentences is a form of entity-based coherence in which the same entity is realized in both sentences, either directly as (69) and (70), or indirectly (71). Note that entity realization here also includes reification of a normal discourse argument mentioned in the first sentence, such as with the demonstrative *this* in (72) and the definite description the appointments in (73).

- (69) [ARG1 *Hale Milgrim, 41 years old, senior vice president, marketing at Elecktra Entertainment Inc., was named president of Capitol Records Inc., a unit of this entertainment concern.*] {EntRel} [ARG2 **Mr. Milgrim succeeds David Berman, who resigned last month.**]
- (70) [ARG1 *The purchase price was disclosed in a preliminary prospectus issued in connection with MGM Grand's planned offering of six million common shares.*] {EntRel} [ARG2 **The luxury airline and casino company, 98.6%-owned by investor Kirk Kerkorian and his Tracinda Corp., earlier this month announced its agreements to acquire the properties, but didn't disclose the purchase price.**]
- (71) [ARG1 *Last year the public was afforded a preview of Ms. Bartlett's creation in a table-model version, at a BPC exhibition.*] {EntRel} [ARG2 **The labels were breathy: "Within its sheltering walls is a microcosm of a thousand years in garden design . . . At the core of it all is a love for plants."**]
- (72) She has done little more than recycle her standard motifs – trees, water, landscape fragments, rudimentary square houses, circles, triangles, rectangles – and fit them into a grid, as if she were making one of her gridded two-dimensional works for a gallery wall. [ARG1 *But for South Gardens, the grid was to be a 3-D network of masonry or hedge walls with real plants inside them.*] {EntRel} [ARG2 **In a letter to the BPCA, Kelly/Varnell called this "arbitrary and amateurish."**]
- (73) [ARG1 *Ronald J. Taylor, 48, was named chairman of this insurance firm's reinsurance brokerage group and its major unit, G.L. Hodson & Son Inc. Robert G. Hodson, 65, retired as chairman but will remain a consultant. Stephen A. Crane, 44, senior vice president and chief financial and planning officer of the parent, was named president and chief executive of the brokerage group and the unit, succeeding Mr. Taylor.*] {EntRel} [ARG2 **The appointments are effective Nov. 1.**]

Entity-based coherence relations are not associated with any sense, their labels being self-evident of their semantic type. Arguments for entity-based relations must be adjacent to each other, though the selection can be discontinuous. The “minimality” constraint here is somewhat restricted, in that the selection should be minimal up to the level of the sentence. In particular, for entity relations we only identify the minimal set of (complete) sentences that mention the entities reified in the Arg2 sentence. Thus, unlike explicit, implicit and alternative lexicalization annotations, arguments of the entity relations cannot comprise a sub-sentential span, including those obtained by excluding attribution. In (74), for instance, the entire sentences are selected as Arg1 and Arg2, even though the “remodeling” and “refurbishing” event entities in Arg1 that are reified and predicated of in Arg2 are embedded as conjoined arguments in the sentential complement, and even though the reification and predication of the same entities in Arg2 should strictly exclude two levels of attribution (see Attribution Section ).

- (74) [ARG1 *Proceeds from the offering are expected to be used for remodeling the company's Desert Inn resort in Las Vegas, refurbishing certain aircraft of the MGM Grand Air unit, and to acquire the property for the new resort.*] {EntRel} [ARG2 **The company said it estimates the**

**Desert Inn remodeling will cost about \$32 million, and the refurbishment of the three DC8-62 aircraft, made by McDonnell Douglas Corp., will cost around \$24.5 million.]**

Example (75) illustrates an annotation of an entity relation where multiple sentence arguments are required. The last sentence only provides an additional predication about the two mentioned ads, but since the antecedent of the referring expression, both ads, is “split” across the previous two sentences, both sentences are selected as Arg1 of the EntRel relation.

- (75) HOLIDAY ADS: Seagram will run two interactive ads in December magazines promoting its Chivas Regal and Crown Royal brands. [ARG1 *The Chivas ad illustrates – via a series of pullouts – the wild reactions from the pool man, gardener and others if not given Chivas for Christmas. The three-page Crown Royal ad features a black-and-white shot of a boring holiday party – and a set of colorful stickers with which readers can dress it up.*] {EntRel} [ARG2 **Both ads were designed by Omnicom’s DDB Needham agency.**]

Supplementary annotations are not allowed for arguments of entity relations. We also do not provide any further annotation within the arguments to identify the entity or entities realized across the arguments: annotation of explicit or implicit anaphoric relations not associated directly with discourse relations is outside the scope of this project.

## Alternative Lexicalization Relations

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

- (76) And she further stunned her listeners by revealing her secret garden design method: [ARG1 *Commissioning a friend to spend “five or six thousand dollars . . . on books that I ultimately cut up.”*] After that{AltLex}, [ARG2 **the layout had been easy**].
- (77) I read the excerpts of Wayne Angell’s exchange with a Gosbank representative (“Put the Soviet Economy on Golden Rails,” editorial page, Oct. 5) with great interest, since the gold standard is one of my areas of research. Mr. Angell is incorrect when he states that the Soviet Union’s large gold reserves would give it “great power to establish credibility.” [ARG1 *During the latter part of the 19th century, Russia was on a gold standard and had gold reserves representing more than 100% of its outstanding currency, but no one outside Russia used rubles. The Bank of England, on the other hand, had gold reserves that averaged about 30% of its outstanding currency, and Bank of England notes were accepted throughout the world.*] The most likely reason for this disparity{AltLex} [ARG2 **is that the Bank of England was a private bank with substantial earning assets, and the common-law rights of creditors to collect claims against the bank were well established in Britain**].
- (78) [ARG1 *Ms. Bartlett’s previous work, which earned her an international reputation in the non-horticultural art world, often took gardens as its nominal subject.*] Mayhap this metaphorical

connection made{AltLex} [ARG2 **the BPC Fine Arts Committee think she had a literal green thumb**].

Annotation of the arguments of alternative lexicalization relations follows the same guidelines as for arguments of implicit connectives. That is, they can be discontinuous and they must include all and only the amount of text minimally required for interpreting the relation.

## “No Relation” Relations

These are cases where no discourse relation or entity-based coherence relation can be inferred between adjacent sentences. The following examples show cases where the “No Relation” label was used.

- (79) The products already available are cross-connect systems, used instead of mazes of wiring to interconnect other telecommunications equipment. [ARG1 *This cuts down greatly on labor, Mr. Buchner said.*] {NoRel} [ARG2 **To be introduced later are a multiplexer, which will allow several signals to travel along a single optical line; a light-wave system, which carries voice channels; and a network controller, which directs data flow through cross-connect systems.**]
- (80) [ARG1 *Jacobs Engineering Group Inc.’s Jacobs International unit was selected to design and build a microcomputer-systems manufacturing plant in County Kildare, Ireland, for Intel Corp. Jacobs is an international engineering and construction concern.*] {NoRel} [ARG2 **Total capital investment at the site could be as much as \$400 million, according to Intel.**]
- (81) While the model was still on view, Manhattan Community Board 1 passed a resolution against South Gardens. The Parks Council wrote the BPCA that this “too ‘private’ . . . exclusive,” complex and expensive “enclosed garden . . . belongs in almost any location but the waterfront.” [ARG1 *Lynden B. Miller, the noted public garden designer who restored Central Park’s Conservatory Garden, recalls her reaction to the South Gardens model in light of the public garden she was designing for 42nd Street’s Bryant Park: “Bryant Park, as designed in 1933, failed as a public space, because it made people feel trapped. By removing the hedges and some walls, the Bryant Park Restoration is opening it up.*] {NoRel} [ARG2 **It seems to me the BPCA plan has the potential of making South Gardens a horticultural jail for people and plants.”]**

As with entity-based elations, “no relation” relations do not imply that the material in Arg2 is not related to anything: It is just that it is not related to the adjacent sentence.

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

## Senses

Explicit, Implicit, and Alternative Lexicalization relations all take one or more sense annotations. These sense annotations fall into four main groups: Temporal, Contingency, Comparison, or Expansion. There is a detailed hierarchy of sense tags, and these are explained in the detailed annotation guide, pages 26-39. Those pages are repeated here for your convenience.

## 4 Senses

### 4.1 Introduction

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

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

### 4.2 Hierarchy of senses

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

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

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

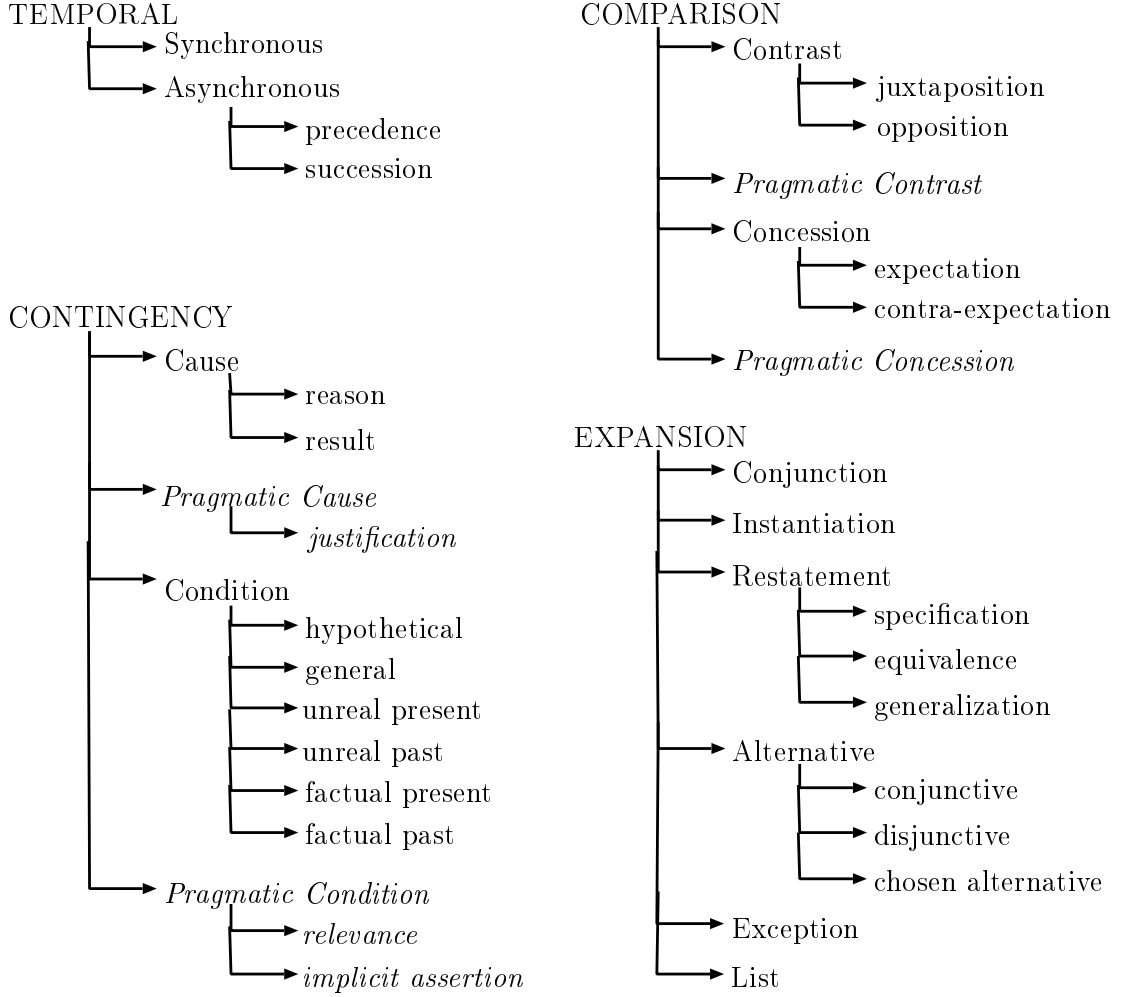


Figure 1: Hierarchy of sense tags

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

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

### 4.3 Class: “TEMPORAL”

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

#### 4.3.1 Type: “Asynchronous”

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

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

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

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

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

#### 4.3.2 Type: “Synchronous”

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

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

### 4.4 Class: “CONTINGENCY”

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

#### 4.4.1 Type: “Cause”

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

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

**Subtype: “reason”.** The type “reason” is used when the connective indicates that the situation described in **Arg2** is the cause and the situation described in **Arg1** is the effect ( $\|Arg2\| < \|Arg1\|$ ), as shown in (102).

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

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

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

#### 4.4.2 Type: “Pragmatic Cause”

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

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

#### 4.4.3 Type: “Condition”

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

---

<sup>17</sup>Logical implication ( $\rightarrow$ ) is used in the rough semantics of “Restatement” (cf. Section 4.6.2).

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

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

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

- (105) Both sides have agreed *that the talks will be most successful* **if negotiators start by focusing on the areas that can be most easily changed**. (CONTINGENCY:Condition:hypothetical) (0082)
- (106) In addition, Black & Decker had said *it would sell two other undisclosed Emhart operations* **if it received the right price**. (CONTINGENCY:Condition:hypothetical) (0807)

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

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

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

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

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

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

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

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

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

- (110) *Of course, **if the film contained dialogue**, Mr. Lane’s Artist would be called a homeless person.* (CONTINGENCY:Condition:unreal present) (0039)
- (111) I’m not saying advertising revenue isn’t important,” she says, “*but I couldn’t sleep at night*” **if the magazine bowed to a company because they once took out an ad.** (CONTINGENCY:Condition:unreal present) (0062)

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

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

#### 4.4.4 Type: “Pragmatic Condition”

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

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

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

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

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

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

## 4.5 Class: COMPARISON

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

### 4.5.1 Type: “Contrast”

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

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

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

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

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

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

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

#### 4.5.2 Type: “Pragmatic Contrast”

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

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

### 4.5.3 Type: “Concession”

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

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

- (119) Although **the purchasing managers’ index continues to indicate a slowing economy, it isn’t signaling an imminent recession**, said Robert Bretz, chairman of the association’s survey committee and director of materials management at Pitney Bowes Inc., Stamford, Conn. (COMPARISON:Concession:expectation) (0036)
- (120) *The Texas oilman has acquired a 26.2% stake valued at more than \$1.2 billion in an automotive-lighting company, Koito Manufacturing Co. But he has failed to gain any influence at the company.* (COMPARISON:Concession:contra-expectation) (0082)
- (121) *Besides, to a large extent, Mr. Jones may already be getting what he wants out of the team, even though it keeps losing.* (COMPARISON:Concession) (1411)

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

## 4.6 Class: “EXPANSION”

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

### 4.6.1 Type: “Instantiation”

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

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

The rough semantics for “Instantiation” involves (1) both arguments holding – ie,  $||\text{Arg1}|| \wedge ||\text{Arg2}||$  – and (2) following (Forbes-Riley *et al.*, 2006), a relation holding between  $||\text{Arg1}||$  and  $||\text{Arg2}||$  of

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

#### 4.6.2 Type: “Restatement”

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

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

- (123) A Lorillard spokeswoman said, “*This is an old story. Implicit = IN FACT **We’re talking about years ago before anyone heard of asbestos having any questionable properties.***” (EXPANSION:Restatement:specification) (0003)
- (124) *An enormous turtle has succeeded where the government has failed: Implicit = SPECIFICALLY **He has made speaking Filipino respectable.*** (EXPANSION:Restatement:specification) (0804)

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

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

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

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

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

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

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

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

#### 4.6.3 Type: “Alternative”

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

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

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

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

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

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

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

---

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

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

#### 4.6.4 Type: “Exception”

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

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

#### 4.6.5 Type: “Conjunction”

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

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

#### 4.6.6 Type: “List”

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

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

### 4.7 Notes on a few connectives

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

#### 4.7.1 Connective: *As if*

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

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

- (136) As if he were still in his old job, Mr. Wright, by resigning with his title instead of being forced from his job, *by law enjoys a \$120,000 annual office expense allowance, three paid staffers, up to \$67,000 for stationery and telephones and continued use of the franking privilege.* (COMPARISON:Concession:contra-expectation) (0909)
- (137) When I realized it was over, I went and stood out in front of the house, waiting and praying for Merrill to come home, *shivering* as if it were 20 below zero until he got there. Never in my life have I been so frightened. (EXPANSION) (1778)

#### 4.7.2 Connective: *Even if*

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

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

#### 4.7.3 Connective: *Otherwise*

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

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

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

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

#### 4.7.4 Connectives: *Or* and *when*

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

- (141) *If you’d really rather have a Buick, don’t leave home without the American Express card. Or so the slogan might go.* (CONTINGENCY:Pragmatic Condition:implicit assertion) (0116)
- (142) *He’s right about his subcommittee’s responsibilities when it comes to obtaining information from prior HUD officials.* (CONTINGENCY:Pragmatic Condition:implicit assertion) (2377)

#### 4.7.5 Connective: *So that*

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

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

## Attribution

Explicit, Implicit, and Alternative Lexicalized relations all take an marking of attribution, which indicates to whom the relation and each argument is attributed. Attribution is covered in the detailed annotation guide, pages 40-49. Those pages are repeated here for your convenience

## 5 Attribution

### 5.1 Introduction

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

- (144) When Mr. Green won a \$240,000 verdict in a land condemnation case against the state in June 1983, he says *Judge O’Kicki unexpectedly awarded him an additional \$100,000.* (0267)
- (145) Advocates said *the 90-cent-an-hour rise, to \$4.25 an hour by April 1991, is too small for the working poor,* while opponents argued **that the increase will still hurt small business and cost many thousands of jobs.** (0098)

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

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

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

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

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

<sup>20</sup>Attribution is also annotated for **AltLex** relations, but not for **EntRel** and **NoRel**, since the latter do not indicate the presence of discourse relations.

- (147) *The gruff financier recently started socializing in upper-class circles.* Implicit = FOR EX-AMPLE Although he says he wasn't keen on going, **last year he attended a New York gala where his daughter made her debut.** (0800)
- (148) *"We've been opposed to" index arbitrage "for a long time,"*  
said Stephen B. Timbers, chief investment officer at Kemper, which manages \$56 billion,  
including \$8 billion of stocks. Implicit = BECAUSE **"Index arbitrage doesn't work, and it scares natural buyers" of stock.** (1000)

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

- (a) *Source*, which distinguishes between different types of agents (Section 5.2);
- (b) *Type*, which encodes the nature of the relationship between agents and AOs, thereby reflecting their factuality (Section 5.3);
- (c) *Scopal polarity*, which is marked when surface negated attribution reverses the polarity of the attributed AO (Section 5.4);
- (d) *Determinacy*, which signals a context that cancels what would otherwise be an entailment of attribution (Section 5.5).

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

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

## 5.2 Source

The *source* feature distinguishes between:

- (a) the writer of the text ("Wr"),
- (b) some specific agent introduced in the text ("Ot" for other),
- (c) some arbitrary ("Arb") individual(s) indicated via a non-specific reference in the text.

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

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

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

	REL	Arg1	Arg2
[Source]	Wr	Inh	Inh

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

	REL	Arg1	Arg2
[Source]	Ot	Inh	Inh

As Example (149) shows, text spans for implicit writer attributions (corresponding to implicit speech acts such as “*I write*”, or “*I say*”) are not marked and imply writer attribution by default.<sup>21</sup>

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

- (151) When Mr. Green won a \$240,000 verdict in a land condemnation case against the State in June 1983, he says *Judge O’Kicki unexpectedly awarded him an additional \$100,000.* (0267)

	REL	Arg1	Arg2
[Source]	Wr	Ot	Inh

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

	REL	Arg1	Arg2
[Source]	Wr	Inh	Ot

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

	REL	Arg1	Arg2
[Source]	Wr	Inh	Ot

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

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

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

	REL	Arg1	Arg2
<b>[Source]</b>	Wr	Ot	Arb

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

	REL	Arg1	Arg2
<b>[Source]</b>	Wr	Inh	Arb

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

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

	REL	Arg1	Arg2
<b>[Source]</b>	Wr	Ot	Ot

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

	REL	Arg1	Arg2
<b>[Source]</b>	Wr	Ot	Ot

---

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

### 5.3 Type

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

#### 5.3.1 Assertion proposition AOs and belief propositions AOs

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

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

- (158) When Mr. Green won a \$240,000 verdict in a land condemnation case against the State in June 1983, he says *Judge O’Kicki unexpectedly awarded him an additional \$100,000.* (0267)

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

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

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

<sup>23</sup>This corresponds roughly to the top-level tier in the AO hierarchy of Asher (1993).

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

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

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

### 5.3.2 Fact AOs

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

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

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

### 5.3.3 Eventuality AOs

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

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

---

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

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

#### 5.4 Scopal polarity

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

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

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

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

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

- (163) a. John doesn't think *Mary will get cured* because **she took the medication**.  
b. John thinks *that* because **Mary took the medication**, *she will not get cured*.

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

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

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

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

## 5.5 Determinacy

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

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

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

## 5.6 Attribution spans

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

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

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

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

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

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

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

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

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

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

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

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

## Differences between this guide and PDTB-2.0

At some point you may need to refer to the original PDTB 2.0 annotation guide to answer a question or clarify a situation. In that case you will need to know how this annotation guide differs from the original guide. Those differences are listed in the following table:

PDTB-2.0	This Guide
Selection of first character of Arg2 to mark textual span of the Implicit, EntRel, and NoRel relation. (p.2)	The first character of Arg2 is not marked; this can be calculated automatically.
The text span for the attribution is annotated as a single (possibly discontinuous) complex reflecting the annotated features, and also includes all non-clausal modifiers of the elements contained in the span. (§5.6)	The text span for each attribution feature (type, source, polarity, and determinacy) is marked individually. If necessary, the full attribution text span may be re-created by combining all the individual feature spans.
While we have annotated modified connectives such as described above, certain types of post-modified connectives have not been annotated, in particular those post-modified by prepositions, for example because (of)..., as a result (of)..., instead (of)..., and rather (than).... While in many cases such expressions relate noun phrases lacking an AO interpretation, (Example 20), there are also a few cases such as Example 21 where they do relate AOs. However, these few tokens have not been annotated. (p. 9-10)	These modified connectives will be annotated the same as other modified connectives.
For practical reasons in the annotation process, all punctuation at the boundaries of connective and argument selections was excluded. (p. 16)	Sentence-ending punctuation is included if the argument contains the whole sentence. Otherwise, argument boundary punctuation is excluded.
Non-clausal attributing phrases are also included obligatorily in the clausal argument they modify (p. 16)	Non-clausal attribution phrases will be treated the same as clausal attribution phrases
While an implicit discourse relation can hold between the final sentence of one paragraph and the initial sentence of the next, implicit relations have not been annotated between adjacent sentences separated by a paragraph boundary (p18)	We will annotate these.
Implicit relations between adjacent clauses in the same sentence not separated by a colon (":") or semi-colon (";") have not been annotated, for example, intra-sentential relations between a main clause and any free adjunct. (p18)	We will annotate these.
We have only annotated implicit relations between adjacent sentences with no Explicit connective between them, even though the presence of an Explicit connective, in particular a discourse adverbial, in a sentence does not preclude the presence of either another Explicit connective relating with the previous text (Example 76) or an Implicit connective (Example 77). (p19)	We will annotate these

the PDTB does not annotate implicit relations between non-adjacent sentences, even if such a relationship holds. (p19)	We will annotate these
--	------------------------

## Glossary

---

(Headwords for entries are in *italics*. A term in a gloss that also appears as a headword in this glossary is underlined.)

*term* – gloss.