# The Class of Discourse Connectives: Open or Closed? Altlex: Alternative Lexicalizations in the Penn Discourse Treebank (PDTB)

Aravind K. Joshi
University of Pennsylvania

Advances in Discourse Analysis and its Computational Aspects
Mumbai, December 15 2012

# What are Discourse Relations?

**Semantically**, discourse relations are well-defined as relations between (two) Abstract Objects, such as events, actions, states, properties, facts/propositions.

**Syntactically**, they can be explicit or implicit

>> **Explicit causal (reason) relation** expressed with a discourse connective:

Increased carbon dioxide emissions will cause the earth to warm up **because** carbon dioxide prevents heat from escaping into space.

>> **Implicit causal (reason) relation** **inferred** between adjacent sentences:

Researchers analyzed changes in concentration of two forms of oxygen. These measurements can indicate temperature changes.

# Penn Discourse Treebank (PDTB)

Annotations of explicit and implicit discourse relations:
- Arguments of discourse relations (Arg1, Arg2)
- Semantics (senses) of discourse relations
- Attribution of discourse relations and their arguments

Corpus: 1 million word Wall Street Journal Corpus
- 2159 texts
- Available through LDC

Project Webpage: http://www.seas.upenn.edu/~pdtb
(provides tools and software for browsing and querying the corpus)

# PDTB Annotation Overview

**Relation Types**

**Explicit Connectives**

**Alternative Lexicalizations (AltLex)**

**Implicit Conn.**

**Entity-based Coherence Relation (EntRel)**
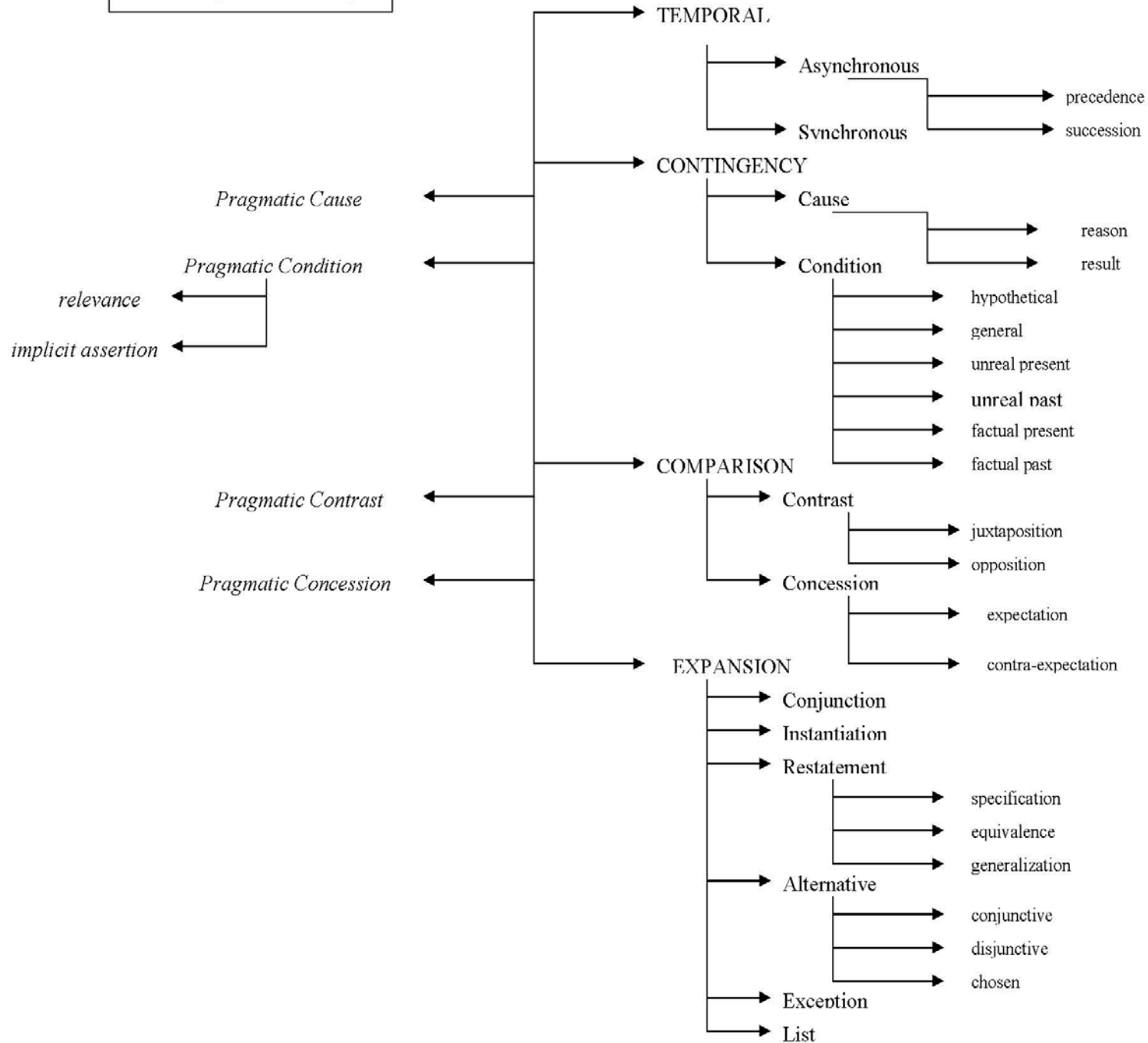
**No Relation (NoRel)**

**Discourse Relations**
(include annotation for semantics and attribution)

**Non-Discourse Relations**
(no annotation for semantics and attribution)

# Hierarchy of sense tags

TEMPORAL
- Asynchronous
  - precedence
  - Synchronous
  - succession

CONTINGENCY
- Cause → Pragmatic Cause
  - reason
  - result
- Condition → Pragmatic Condition
  - relevance
  - implicit assertion
  - hypothetical
  - general
  - unreal present
  - unreal past
  - factual present
  - factual past

COMPARISON
- Contrast → Pragmatic Contrast
  - juxtaposition
  - opposition
- Concession → Pragmatic Concession
  - expectation
  - contra-expectation

EXPANSION
- Conjunction
- Instantiation
- Restatement
  - specification
  - equivalence
  - generalization
- Alternative
  - conjunctive
  - disjunctive
  - chosen
- Exception
- List

# Lexicalization of Discourse Relations

A common belief (Quirk et al. (1972), Knott (1996)) is that explicitly realized discourse relations (explicit connectives) can be defined as belonging to **well-defined syntactic classes** and that they are further **closed class items**.

(1) Subordinating conjunctions (because, when, although, etc)

(2) Coordinating conjunctions (and, but, or, .)

(3) Discourse adverbials (however, as a result, for example, etc.)

(4) Prepositional phrases containing propositional anaphora referring
     back to one of the abstract object arguments
     (after that, despite that, etc.)

(5) Phrases that take sentential complements (this means, that's why,
     the fact is that, etc.)

# Explicit Discourse Connectives in PDTB

PDTB annotation of explicit connectives started with the closed-class conception, but with only three syntactic classes
- subordinating conjunctions
- coordinating conjunctions
- discourse adverbials


Annotation procedure:

      (1) Identify and mark the explicit discourse **connectives**
      (2) Identify and mark their **arguments**
      (3) Label **sense** or senses of connective


Lists of discourse connectives were provided to annotators.

# Explicit Discourse Connectives: Examples

>> *The federal government suspended sales of U.S. savings bonds* because (cause-reason) Congress hasn't lifted the ceiling on government debt. [wsj_0008]

>> *The subject will be written into the plots of prime-time shows*, and (conjunction) viewers will be given a 900 number to call. [wsj_2100]

>> *In the past, the socialist policies of the government strictly limited the size of … industrial concerns to conserve resources and restrict the profits businessmen could make.*
As a result (cause-result), industry operated out of small, expensive, highly inefficient industrial units. [wsj_0629]

**Arg2** is syntactically associated with the connective.
*Arg1* is the other argument (Arg1 can be distant)

# Implicit Discourse Connectives in PDTB

In adjacent sentence contexts, we also annotated implicit connectives when there were no "explicit connectives" to relate the two sentences and when the relation had to be **inferred** by the annotator.

Annotation procedure:

    (1) Identify the relation **inferred** between the sentences
    (2) **Insert** a connective that best expresses the relation and sounds fluent.
    (3) Label the **sense** of inferred relation

# Implicit Discourse Connectives: Examples

>> *Some have raised their cash positions to record levels.* Implicit=because (cause-reason) High cash positions help buffer a fund when the market falls. [wsj_0983]

>> *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 (cause-result) By a rule of thumb of 1.5 new jobs for each new hotel room, Clark County will have nearly 18,000 new jobs. [wsj_0994]

Arg2 is the second sentence.
*Arg1* is the first sentence.

# When Implicit Connectives could NOT be Inserted

In annotating implicit relations in these adjacent sentence contexts, annotators were not able to insert connectives in many cases! They inserted "NONE" as the connective.
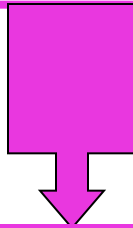
In a later phase of the annotation, NONE tokens (approx. 6000) were analyzed further.

About 10% of these tokens did in fact express a discourse relation, but annotators were nevertheless unable to insert a connective!

# When Implicit Connectives could NOT be Inserted

Annotators were unable to insert a connective despite the inference of a discourse relation because there was a perceived redundancy after insertion of the connective, and thus the connective did not meet the fluency criteria.

**Source of Redundancy**: Discourse relation was realized by an expression that had not been pre-classified as a discourse connective.

**Alternative Lexicalizations (AltLex)**

# Examples of AltLex in PDTB

| AltLex Expression | (Syntax) | Sense | Adv.Connective Counterpart |
|---|---|---|---|
| Trouble is | (NP-SBJ V) | Concession | However |
| At the other end of the spectrum | (PP-LOC) | Contrast | ?? |
| The reason: | (NP) | Cause-Reason | NONE? |
| That means | (NP-SBJ V) | Cause-Result | As a result |
| Beyond that | (PP) | Conjunction | In addition |
| Probably the most egregious example | (ADVP NP-SBJ V) | Instantiation | ?? |
| Putting it all Together | (S-ADV) | Restatement | In sum |
| That was followed by | (NP-SBJ V V P) | Temporal | Then |

# AltLex Analysis: Some are Somewhat Closed-class

Some AltLex's are somewhat closed class expressions and potential
    connectives once propositional anaphoric pronouns referring to
    Arg1 are also allowed to be part of the connective phrase.
    Examples from Knott (1996):
    - After that, after this,
    - That's why, that is why, this is why,
    - This means, that means

 We have found many new items that don't appear in previous lists:
        - trouble (with that) is, the idea (behind that) is,
        the problem (regarding that) is, the reason (for that) is,
        the result (of that) is, etc.

Indeed, many attested adverbial connectives are argued to have
implicit propositional anaphora (Forbes, 2003)

# Closed-class AltLex: Examples

>> *Certainly, the Oct. 13 sell-off didn't settle any stomachs*. **Beyond that (conjunction)**, **money managers and analysts see other problems.** [wsj_0359]

>> *She spent a month at an Aetna school in Gettysburg, Pa., learning all about the construction trade, including masonry, plumbing and electrical wiring*. **That was followed by (temporal)** **three months at the Aetna Institute in Hartford, where she was immersed in learning how to read and interpret policies.** [wsj_0766]

>> *Mr. Payson, an art dealer and collector, sold Vincent van Gogh's "Irises" at a Sotheby's auction in November 1987 to Australian businessman Alan Bond*. **Trouble is (Concession)**, **Mr. Bond has yet to pay up, and until he does, Sotheby's has the painting under lock and key.** [wsj_2113]

# Closed-class AltLex: Examples

>> *In addition, Unisys must deal with its increasingly oppressive debt load. Debt has risen to around $4 billion, or about 50% of total capitalization*. **That means (cause-result)** **Unisys must pay about $100 million in interest every quarter, on top of $27 million in dividends on preferred stock**. [wsj_0568]

>> *Both are in great need of foreign exchange, and South Africa is also under pressure to meet foreign loan commitments*, he said. **"Putting it all together (restatement)**, **we have a negative scenario that doesn't look like it will improve overnight**," he said. [wsj_1687]

**Cause-Reason** is the only listed sense for which there are no attested adverbial counterparts in English. The preferred way to realize this relation inter-sententially is as an AltLex.

>> *After trading at an average discount of more than 20% in late 1987 and part of last year, country funds currently trade at an average premium of 6%.* **The reason: (cause-reason)** **Share prices of many of these funds this year have climbed much more sharply than the foreign stocks they hold.** [wsj_0034]

Is this specific to English, or a linguistic universal?

Hindi, Czech, Turkish, Italian, Arabic

# AltLex Analysis: Cause-Reason is Special

There are 7 out of 858 instances of **because** observed in PDTB that appear as adverbs.

4 of these are in QA contexts

> >> *"Why was containment so successful?*
> **Because** it had bipartisan support." [wsj_0771]

# 'Because' as an Adverb?

Are the remaining 3 simply stylistic aberrations, or evidence of **because** emerging as an adverb as well?

>> Many of us are suckers. *But what we may not know is just what makes somebody a sucker*. What makes people blurt out their credit-card numbers to a caller they've never heard of? Do they really believe that the number just for verification and is simply a formality on the road to being a grand-prize winner? What makes a person buy an oil well from some stranger knocking on the screen door? Or an interest in a retirement community in Nevada that will knock your socks off, once it is built?

**Because in the end, these people always wind up asking themselves the same question: "How could I be so stupid?"** [wsj_1572]

N.B: There is other evidence of connectives behaving similarly, e.g., so, but, and (but their adverbial use alternates with their use as coordinating, not subordinating conjunction.

# 'Because' as an Adverb?

>> Players ran out on the field way below, and the stands began to reverberate. It must be a local custom, I thought, stamping feet to welcome the team. But then the noise turned into a roar.
And no one was shouting. *No one around me was saying anything*. **Because** **we all were busy riding a wave**. Sixty thousand surfers atop a concrete wall, waiting for the wipeout. [wsj_1643]

>> President Bush told reporters: "Whether that {the leadership change} reflects a change in East-West relations, *I don't think so*. **Because** **Mr. Krenz has been very much in accord with the policies of Honecker**." [wsj_1875]

# AltLex Analysis: Not so closed-class

>> *Inflation is expected to be highest in Greece, where it is projected at 14.25%, and Portugal, at 13%.*
**At the other end of the spectrum (contrast)**, West German inflation was forecast at 3% in 1989 and 2.75% in 1990.

>> *Typically, these laws seek to prevent executive branch officials from inquiring into whether certain federal programs make any economic sense or proposing more market-oriented alternatives to regulations.* **Probably the most egregious example is (instantiation)** a proviso in the appropriations bill for the executive office that prevents the president's Office of Management and Budget from subjecting agricultural marketing orders to any cost-benefit scrutiny. There is something inherently suspect about Congress's prohibiting the executive from even studying whether public funds are being wasted in some favored program or other. [wsj_0112]

# When do the open-ended AltLex's Occur?

**Relation modification to convey more than the bare connective can convey**

We do have "modified connectives": e.g., possibly because

But many adverbial connective forms do not allow
Modification - * possibly for example

Modification is possible only after Altlexification!
- a possible example (NP)

-Eventually, some of these may get grammaticized in much the same manner as some current day adverbials - cf. therefore

# AltLex Analysis: Not so closed-class

>> *Inflation is expected to be highest in Greece, where it is projected at 14.25%, and Portugal, at 13%.*
**At the other end of the spectrum (contrast)**, **West German inflation was forecast at 3% in 1989 and 2.75% in 1990.**

>> *Typically, these laws seek to prevent executive branch officials from inquiring into whether certain federal programs make any economic sense or proposing more market-oriented alternatives to regulations.* **Probably the most egregious example is (instantiation)** **a proviso in the appropriations bill for the executive office that prevents the president's Office of Management and Budget from subjecting agricultural marketing orders to any cost-benefit scrutiny. There is something inherently suspect about Congress's prohibiting the executive from even studying whether public funds are being wasted in some favored program or other.** [wsj_0112]

# Summary

- PDTB annotations of connectives and their arguments
    -- explicit and implicit connectives
    -- Altlex

- Discourse Connectives: Open or Closed class
    -- Explicit: closed
    -- Altlex: Open or Closed or CLOPEN?
-         - Partly open
        - Why are there Altlex items?
        -  Impossible Altlex?
            -- Impossible adverbial Altlex with the
                sense "cause-reason" ?
                --- Is this  a universal?

AltLex's in other languages:

Czech : Magdalena Rysova
Hindi: Hyderabad, IIIT Group- Dipti Sharma
Turkish: METU Group
.
.
.