Brandeis University

Multilingual Discourse Annotation

Nianwen Xue

7/19/2011

LSA Summer Institute

Acknowledgement: many slides provided by Aravind Joshi

Outline

- Role of Annotated Corpora at the discourse level
  - Moving to annotations at the discourse level
- A brief description of the Penn Discourse Treebank (PDTB)
  - Annotations of explicit and implicit connectives and their arguments
  - Attributions
  - Senses of connectives
- Comparison with complexity of dependencies at the sentence level
- Summary

What is a discourse relation?

The meaning and coherence of a discourse results partly from how its constituents relate to each other.

- **Reference relations**
- **Discourse relations**
  - **Informational**
  - **Intentional**

Informational discourse relations convey relations that hold in the subject matter.

Intentional discourse relations specify how intended discourse effects relate to each other.

Why Discourse Relations?

Discourse relations provide a level of description that is

- theoretically interesting, linking sentences (clauses) and discourse
- identifiable more or less reliably on a sufficiently large scale
- capable of supporting a level of inference potentially relevant to many NLP applications.

Discourse Annotation Resources

- RST Discourse Treebank
  - Based on Rhetorical Structure Theory (Mann and Thompson, 1988)
- Discourse Graphbank
- Penn Discourse Treebank
  - Based on Discourse Lexicalized TAG (Webber, Joshi, Stone, Knott, 2003)

Basic research questions

- What is the nature of discourse relations?
  - Conceptual relations between abstract objects
  - Lexically grounded relations?
- What is the inventory of discourse relations?
- What is the appropriate data structure for discourse relations
  - Trees
  - Graphs
  - Dependencies

RST answers

- What is the nature of discourse relations?
  - Conceptual relations between abstract objects
  - ~~Lexically grounded relations?~~
- What is the inventory of discourse relations?
  - See RST Corpus annotation manual
- What is the appropriate data structure for discourse relations
  - Trees
  - ~~Graphs~~
  - ~~Dependencies~~

RST data structure

- Discourse structure modeled by schemas (expressed as context-free rules)
- Leaves are an elementary discourse units (a continuous text span)
- Non-terminals cover contiguous, non-overlapping text spans
- Discourse relations (aka rhetorical relations) hold only between daughters of the same non-terminal

PDTB answers

- What is the nature of discourse relations?
  - ~~Conceptual relations between abstract objects~~
  - Lexically grounded relations
- What is the inventory of discourse relations?
  - See PDTB sense hierarchy
- What is the appropriate data structure for discourse relations
  - Structures and dependencies
  - Does not assume tree structure *a priori*

Operational decisions

- Lexically grounded approach
- Adjacent sentences
- Arg1 and arg2 conveniently defined
  - Only 2 AO arguments, labeled ***Arg1*** and **Arg2**
  - **Arg2**: clause with which connective is syntactically associated
  - ***Arg1***: the other argument
- No comma delimited discourse relations

How are Discourse Relations triggered in PDTB?

Lexical Elements and Structure

- **Lexically-triggered discourse relations can relate the Abstract Object interpretations of non-adjacent as well as adjacent components. Discourse connectives serve as the lexical triggers**

- **Discourse relations can be triggered by structure underlying adjacency, i.e., between adjacent components unrelated by lexical elements.**

Lexical approach to syntax at the sentence level pushed up to discourse

Sources of discourse meaning resemble the sources of sentence meaning, for example,

- **structure:** **e.g., verbs and their arguments conveying pred-arg relations;**
- **adjacency:** **e.g., noun-noun modifiers conveying relations implicitly;**
- **anaphora:** **e.g., modifiers like *other* and *next*, conveying relations anaphorically.**

Lexical Triggers

Discourse connectives (explicit):

- coordinating conjunctions
- subordinating conjunctions and subordinators
- paired (parallel) constructions
- discourse adverbials
- Others

Discourse connectives (implicit): Introduced, when appropriate, between adjacent sentences when no explicit connectives are present

Penn Discourse Treebank (PDTB)

- Wall Street Journal (same as the Pen Treebank (PTB) corpus): ~1M words
- Annotation record
  -- the text spans of connectives and their arguments
  -- features encoding the semantic classification of connectives, and attribution of connectives and their arguments.
- PDTB 1.0 (April 2006), PDTB 2.0 (January 2008), through LDC) PDTB Project: UPENN: Nikhil Dinesh, Aravind Joshi, Alan Lee, Eleni Miltsakai, Rashmi Prasad, and U. Edinburgh: Bonnie Webber (supported by NSF)
- http://www.seas.upenn.edu/~pdtb
  -- Documentation of Annotation Guidelines, papers, tutorials, tools, link to LDC

Explicit Connectives

Explicit connectives are the lexical items that trigger discourse relations.

- Subordinating conjunctions (e.g., *when*, *because*, *although,* etc.)
  - ***The federal government suspended sales of U.S. savings bonds*** **because** **Congress hasn't lifted the ceiling on government debt**.
- Coordinating conjunctions (e.g., *and*, *or*, *so*, *nor*, etc.)
  - ***The subject will be written into the plots of prime-time shows***, **and** **viewers will be given a 900 number to call**.
- Discourse adverbials (e.g., *then*, *however*, *as a result*, etc.)
  - ***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**, **industry operated out of small, expensive, highly inefficient industrial units**.

Identifying Explicit Connectives

Primary criterion for filtering: Arguments must denote Abstract Objects.

The following are rejected because the AO criterion is not met

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

- 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.

Modified Connectives

Connectives can be modified by adverbs and focus particles:

- ➢ *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.

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

- ▪ Initially identified connective (since, if) is extended to include modifiers.
- ☞ Each annotation token includes both head and modifier (e.g., even if).
- ☞ Each token has its head as a feature (e.g., if)

Parallel Connectives

Paired connectives take the same arguments:

- On the one hand, Mr. Front says, *it would be misguided to sell into "a classic panic*." On the other hand, it's not necessarily a good time to jump in and buy.
- Either *sign new long-term commitments to buy future episodes* or risk losing "Cosby" to a competitor.

- Treated as complex connectives – annotated discontinuously
- Listed as distinct types (no head-modifier relation)

Complex Connectives

Multiple relations can sometimes be expressed as a conjunction of connectives:

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

- Treated as complex connectives
- Listed as distinct types (no head-modifier relation)

Argument Labels and Linear Order

- **Arg2 is the sentence/clause with which connective is syntactically associated. *Arg1* is the other argument.**

- **No constraints on relative order. Discontinuous annotation is allowed.**

  - **Linear:**
    - ***The federal government suspended sales of U.S. savings bonds*** **because Congress hasn't lifted the ceiling on government debt.**

  - **Interposed:**
    - ***Most oil companies*****, when they set exploration and production budgets for this year, *forecast revenue of $15 for each barrel of crude produced*.**

    - ***The chief culprits*****, he says, *are big companies and business groups that buy huge amounts of land "not for their corporate use, but for resale at huge profit*." … The Ministry of Finance, as a result, has proposed a series of measures that would restrict business investment in real estate even more tightly than restrictions aimed at individuals.**

Location of Arg1

- **Same sentence as Arg2:**
  - ***The federal government suspended sales of U.S. savings bonds*** **because** **Congress hasn't lifted the ceiling on government debt**.
- **Sentence immediately previous to Arg2:**
  - ***Why do local real-estate markets overreact to regional economic cycles?*** **Because** **real-estate purchases and leases are such major long-term commitments that most companies and individuals make these decisions only when confident of future economic stability and growth**.
- **Previous sentence non-contiguous to Arg2 :**
  - Mr. Robinson … said ***Plant Genetic's success in creating genetically engineered male steriles doesn't automatically mean it would be simple to create hybrids in all crops***. That's because pollination, while easy in corn because the carrier is wind, is more complex and involves insects as carriers in crops such as cotton. "It's one thing to say you can sterilize, and another to then successfully pollinate the plant," he said. **Nevertheless**, he said, **he is negotiating with Plant Genetic to acquire the technology to try breeding hybrid cotton**.

Types of Arguments

- Simplest syntactic realization of an Abstract Object argument is:
  - A **clause**, tensed or non-tensed, or ellipsed.

  The clause can be a matrix, complement, coordinate, or subordinate clause.

- A Chemical spokeswoman said ***the second-quarter charge was "not material"*** **and that no personnel changes were made** **as a result**.
- In Washington, House aides said Mr. Phelan told congressmen that the collar, ***which banned program trades through the Big Board's computer*** **when** **the Dow Jones Industrial Average moved 50 points**, didn't work well.
- ***Knowing a tasty -- and free -- meal*** **when** **they eat one**, the executives gave the chefs a standing ovation.

☞ **Syntactically implicit elements for non-finite and extracted clauses are assumed to be available.**

- ***Players for the Tokyo Giants, for example, must always wear ties*** **when** **on the road.**

Multiple Clauses: Minimality Principle

- Any number of clauses can be selected as arguments:
  - *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, 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.

But, the selection is constrained by a Minimality Principle:

- Only as many clauses and/or sentences should be included as are minimally required for interpreting the relation. Any other span of text that is perceived to be relevant (but not necessary) should be annotated as supplementary information:
  - Sup1 for material supplementary to *Arg1*
  - Sup2 for material supplementary to Arg2

Conventions

- Discontinuous annotation is allowed when including non-clausal modifiers and heads:

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

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

Annotation Overview: Explicit Connectives

- All WSJ sections (25 sections; 2304 texts)

- 100 distinct types
  - Subordinating conjunctions – 31 types
  - Coordinating conjunctions – 7 types
  - Discourse Adverbials – 62 types

  (Some additional types are annotated for PDTB-2.0.)

- About 20,000 distinct tokens

Implicit Connectives

When there is no Explicit connective present to relate adjacent sentences, it may be possible to infer a discourse relation between them due to adjacency.

- *Some have raised their cash positions to record levels*. Implicit=? High cash positions help buffer a fund when the market falls.
- *The projects already under construction will increase Las Vegas's supply of hotel rooms by 11,795, or nearly 20%, to 75,500*. Implicit=?) 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 implicit connectives are annotated by inserting a connective that "best" captures the relation.

- Sentence delimiters are: period, semi-colon, colon
- Left character offset of Arg2 is "placeholder" for these implicit connectives.

Implicit Connectives

When there is no Explicit connective present to relate adjacent sentences, it may be possible to infer a discourse relation between them due to adjacency.

- *Some have raised their cash positions to record levels*. Implicit=because (causal) High cash positions help buffer a fund when the market falls.

- *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 (consequence) 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 implicit connectives are annotated by inserting a connective that "best" captures the relation.

- Sentence delimiters are: period, semi-colon, colon
- Left character offset of Arg2 is "placeholder" for these implicit connectives.

Where Implicit Connectives are Not Annotated

- Intra-sententially, e.g., between main clause and free adjunct:
  - (Consequence: so/thereby) Second, they channel monthly mortgage payments into semiannual payments, reducing the administrative burden on investors.
  - (Continuation: then) Mr. Cathcart says he has had "a lot of fun" at Kidder, adding the crack about his being a "tool-and-die man" never bothered him.
- Implicit connectives in addition to explicit connectives: If at least one connective appears explicitly, any additional ones are not annotated:
  - (Consequence: so) On a level site you can provide a cross pitch to the entire slab *by raising one side of the form*, but for a 20-foot-wide drive this results in an awkward 5-inch slant. Instead, make the drive higher at the center.

Decision point 4:

Extent of Arguments of Implicit Connectives

- **Like the arguments of Explicit connectives, arguments of Implicit connectives can be sentential, sub-sentential, multi-clausal or multi-sentential:**
  - **Legal controversies in America have a way of assuming a symbolic significance far exceeding what is involved in the particular case. They speak volumes about the state of our society at a given moment. *It has always been so*. Implicit=for example (exemplification) 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.**

Non-insertability of Implicit Connectives

There are three types of cases where Implicit connectives cannot be inserted between adjacent sentences.

- AltLex: A discourse relation is inferred, but insertion of an Implicit connective leads to redundancy because the relation is Alternatively Lexicalized by some non-connective expression:
  - *Ms. Bartlett's previous work, which earned her an international reputation in the non-horticultural art world, often took gardens as its nominal subject*. AltLex = (consequence) **Mayhap this metaphorical connection made the BPC Fine Arts Committee think she had a literal green thumb**.

Non-insertability of Implicit Connectives

- **EntRel: the coherence is due to an entity-based relation.**
  - *Hale Milgrim, 41 years old, senior vice president, marketing at Elecktra Entertainment Inc., was named president of Capitol Records Inc., a unit of this entertainment concern*. EntRel Mr. Milgrim succeeds David Berman, who resigned last month.
- **NoRel: Neither discourse nor entity-based relation is inferred.**
  - *Jacobs is an international engineering and construction concern*. NoRel Total capital investment at the site could be as much as $400 million, according to Intel.

☞ **Since EntRel and NoRel do not express discourse relations, no semantic classification is provided for them.**

Annotation overview: Implicit Connectives

- About 18,000 tokens
  - **Implicit Connectives**: about 14,000 tokens
  - **AltLex**: about 200 tokens
  - **EntRel**: about 3200 tokens
  - **NoRel**: about 350 tokens

Annotation Overview: Attribution

- Attribution features are annotated for
  - Explicit connectives
  - Implicit connectives
  - AltLex

☞ **34% of discourse relations are attributed to an agent other than the writer.**

Attribution

Attribution captures the relation of “ownership” between agents and Abstract Objects.

☞ But it is not a discourse relation!

Attribution is annotated in the PDTB to capture:

(1) How discourse relations and their arguments can be *attributed to different individuals*:

- 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*.

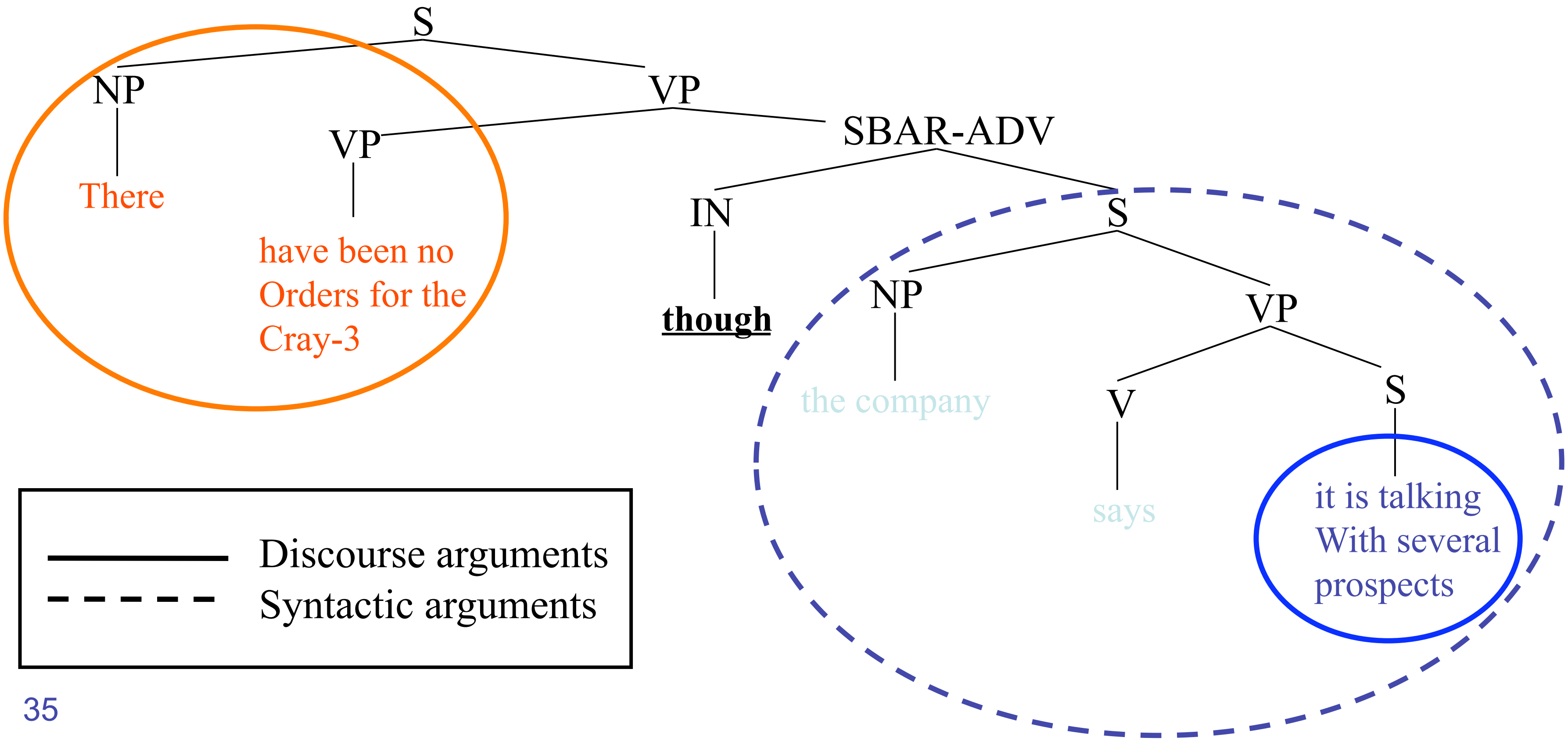
⇒ Relation and Arg2 are attributed to the Writer.

⇒ *Arg1* is attributed to another agent.

- There have been no orders for the Cray-3 so far, **though** the company says it is talking with several prospects.

✓ Discourse semantics: contrary-to-expectation relation between “there being no orders for the Cray-3” and “there being a possibility of some prospects”.

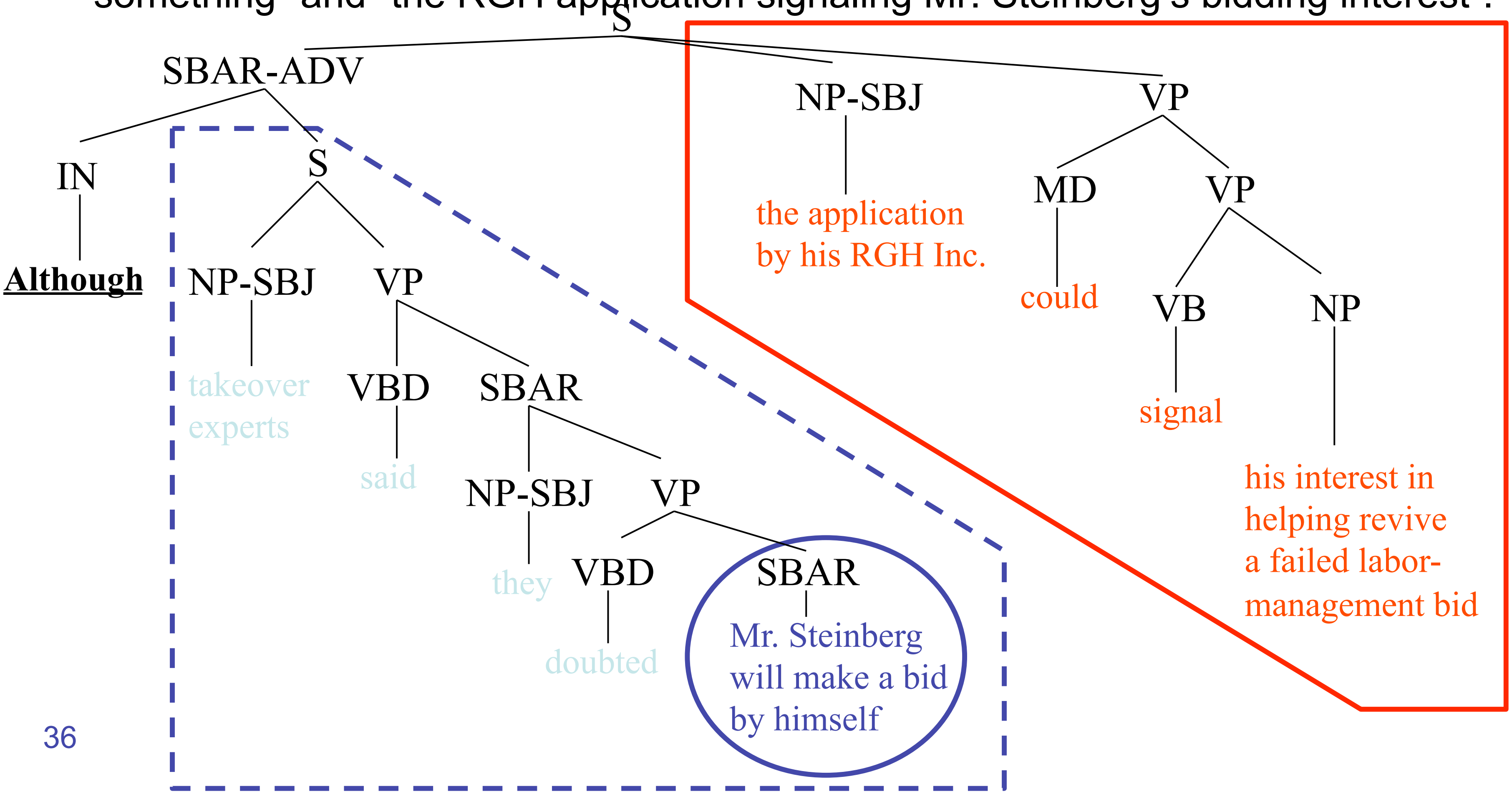
✖ Sentence semantics: contrary-to-expectation relation between “there being no orders for the Cray-3” and “the company saying something”.



- **Although** takeover experts said they doubted Mr. Steinberg will make a bid by himself, the application by his Reliance Group Holdings Inc. could signal his interest in helping revive a failed labor-management bid.

✓ Discourse semantics: contrary-to-expectation relation between "Mr. Steinberg not making a bid by himself" and "the RGH application signaling his bidding interest".

✖ Sentence semantics: contrary-to-expectation relation between "experts saying something" and "the RGH application signaling Mr. Steinberg's bidding interest".


S
SBAR-ADV
IN
Although
S
NP-SBJ
takeover experts
VP
VBD
said
SBAR
NP-SBJ
they
VP
VBD
doubted
SBAR
Mr. Steinberg will make a bid by himself
NP-SBJ
the application by his RGH Inc.
VP
MD
could
VP
VB
signal
NP
his interest in helping revive a failed labor-management bid

- Mismatches occur with other relations as well, such as causal relations:

➢ Credit analysts said investors are nervous about the issue <u>**because**</u> they say the company's ability to meet debt payments is dependent on too many variables, including the sale of assets and the need to mortgage property to retire some existing debt.

✓ Discourse semantics: causal relation between “investors being nervous” and “problems with the company’s ability to meet debt payments”

✖ Sentence semantics: causal relation between “investors being nervous” and “credit analysts saying something”!

- Attribution cannot always be excluded by default

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

Attribution Features

Attribution is annotated on relations and arguments, with FOUR features

- **Source: encodes the different agents to whom proposition is attributed**
  - **Wr: Writer agent**
  - **Ot: Other non-writer agent**
  - **Arb: Generic/Atbitrary non-writer agent**
  - **Inh: Used only for arguments; attribution inherited from relation**
- **Type: encodes different types of Abstract Objects**
  - **Comm: Verbs of communication**
  - **PAtt: Verbs of propositional attitude**
  - **Ftv: Factive verbs**
  - **Ctrl: Control verbs**
  - **Null: Used only for arguments with no explicit attribution**

Attribution Features (cont'd)

- Polarity: encodes when surface negated attribution interpreted lower
  - Neg: Lowering negation
  - Null: No Lowering of negation

- Determinacy: indicates that the annotated TYPE of the attribution relation cannot be taken to hold in context
  - Indet: is used when the context cancels the entailment of attribution
  - Null: Used when no such embedding contexts are present

Annotations of Senses of Connectives in PDTB

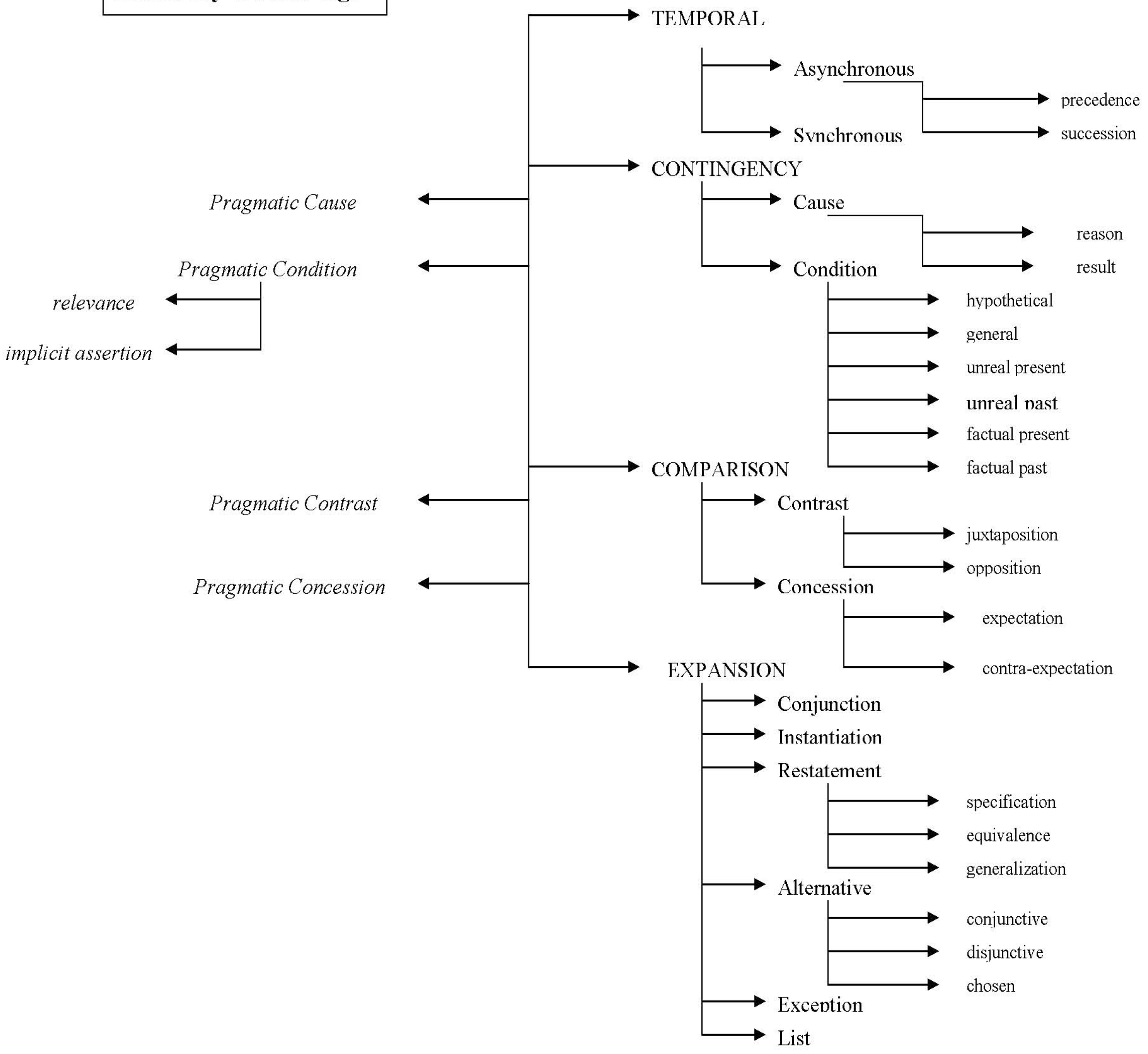
- Sense annotations for explicit, implicit and altlex tokens

- Total: 35,312 tokens

Annotation and adjudication

- Predefined sets of sense tags
- 2 annotators
- Adjudication
  - Agreeing tokens → No adjudication
  - Disagreement at third level (subtype) → second level tag (type)
  - -Disagreement at second level (type) → first level tag (class)
  - Disagreement at class level →adjudicated

Hierarchy of sense tags



Sense Tags

Sense tags are organized hierarchically

- A CLASS level tag is mandatory
- The Type level provides a more specific interpretation of the relation between the situations described in Arg1 & Arg2
- The subtype level describes the specific contribution of the arguments to the interpretation of the relation (e.g. which situation is the cause and which is the result)
- Types and subtypes are optional: They apply when the annotators can comfortably identify a finer or more specific interpretation
- A Type or CLASS level tag also applies when the relation between arg1 and arg2 is ambiguous between two finer interpretations (e.g. COMPARISON may apply when both a contrastive and a concessive interpretations are available)

First level: CLASSES

- Four CLASSES
  - TEMPORAL
  - CONTINGENCY
  - COMPARISON
  - EXPANSION

Second level: Types

- TEMPORAL
  - Asynchronous
  - Synchronous
- CONTINGENCY
  - Cause
  - Condition
- COMPARISON
  - Contrast
  - Concession
- EXPANSION
  - Conjunction
  - Instantiation
  - Restatement
  - Alternative
  - Exception
  - List

Third level: subtype

- TEMPORAL: Asynchronous
  - Precedence
  - Succession
- TEMPORAL: Synchronous

  No subtypes

- CONTINGENCY: Cause
  - reason
  - Result
- CONTINGENCY: Condition
  - hypothetical
  - general
  - factual present
  - factual past
  - unreal present
  - unreal past

Third level: subtype

- COMPARISON: Contrast
  - Juxtaposition
  - Opposition
- COMPARISON: Concession
  - expectation
  - contra-expectation
- EXPANSION: Restatement
  - Specification
  - Equivalence
  - Generalization
- EXPANSION: Alternative
  - Conjunctive
  - Disjunctive
  - Chosen alternative

Semantics of CLASSES

- TEMPORAL
  - The situations described in Arg1 and Arg2 are temporally related
- CONTINGENCY
  - The situations described in Arg1 and Arg2 are causally influenced
- COMPARISON
  - The situations described in Arg1 and Arg2 are compared and *differences* between them are identified *(similar situations do not fall under this CLASS)*
- EXPANSION
  - The situation described in Arg2 provides information deemed relevant to the situation described in Arg1

Semantics of Types/subtypes

- TEMPORAL: Asynchronous: temporally ordered events
  - precedence: Arg1 event precedes Arg2
  - succession: Arg1 event succeeds Arg1
- TEMPORAL: Synchronous: temporally overlapping events
- CONTINGECY: Cause: events are causally related
  - Reason: Arg2 is cause of Arg1
  - Result: Arg2 results from Arg1
- CONTINGENCY: Condition: if Arg1 → Arg2
  - Hypothetical: Arg1 → Arg2 (evaluated in present/future)
  - General: everytime Arg1 → Arg2
  - Factual present: Arg1 → Arg2 & Arg1 taken to hold at present
  - Factual past: Arg1 →Arg2 & Arg1 taken to have held in past
  - Unreal present: Arg1→ Arg2 & Arg1 is taken not to hold at present
  - Unreal past: Arg1 → Arg2 & Arg1 did not hold → Arg2 did not hold

- COMPARISON: Contrast: differing values assigned to some aspect(s) of situations described in Arg1&Arg2
  - Juxtaposition: specific values assigned from a range of possible values (e.g.,
  - Opposition: antithetical values assigned in cases when only two values are possible
- COMPARISON: Concession: expectation based on one situation is denied
  - Expectation: Arg2 creates an expectation C, Arg1 denies it
  - Contra-expectation: Arg2 denies an expectation created in Arg1

- EXPANSION
  - Conjunction: additional discourse new information
  - Instantiation: Arg2 is an example of some aspect of Arg1
  - Restatement: Arg2 is about the same situation described in Arg1
    - Specification: Arg2 gives more details about Arg1
    - Equivalence: Arg2 describes Arg1 from a different point of view
    - Generalization: Arg2 gives a more general description/conclusion of the situation described in Arg1
  - Alternative: Arg1&Arg2 evoke alternatives
    - Conjunctive: both alternatives are possible
    - Disjunctive: only one alternative is possible
    - Chosen alternative: two alternative are evoked, one is chosen (semantics of “instead”)
  - Exception: Arg1 would hold if Arg2 didn’t
  - List: Arg1 and Arg2 are members of a list

Pragmatic tags

- Pragmatic cause: justification
  - Mrs Yeargin is lying. (BECAUSE) They found students in an advanced class a year earlier who said she gave them similar help

- Pragmatic condition: relevance, implicit assertion
  - Rep. John Dingell is trying again to raise the Fairness Doctrine from the dead if the White House is looking for another unconstitutional bill (relevance)
  - If any nation can use environmentally benign architecture, it is Poland. (implicit assertion)

- Pragmatic contrast: contrast between some situation/evaluation inferred on the basis of Arg1
  - 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 (infer “but that’s not the only reason”)

Examples

- EXPANSION: Instantiation
  - In some respects they [hypertext books] are clearly superior to normal books, for example they have database cross-referencing facilities ordinary volumes lack
- EXPANSION: Restatement: generalization
  - John has given his sister a lot of money, then he helped his kid in doing homeworks and finally he washed my car*.* In sum, John is a very good man.
- EXPANSION: Restatement: equivalence
  - 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: Exception
  - Boston Co. officials declined to comment 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.

Patterns of Dependencies in the PDTB

- Connectives and their arguments have been annotated individually and independently

- What patterns do we find in the PDTB with respect to pairs of consecutive connectives?

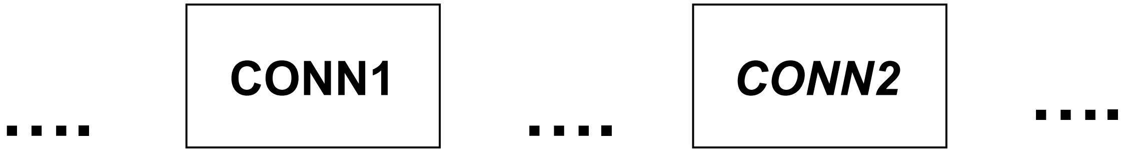
- The annotations does not necessarily lead to a single tree over the entire discourse

  -- comparison with the sentence level

- Complexity of discourse dependencies?

  -- comparison with the sentence level.

Patterns of Consecutive Connectives



…. CONN1 …. *CONN2* ….

1. How do the text spans associated with Conn1 and its args relate to those of Conn2 and its args?

2. Do the pred-arg dependencies of Conn1 cross those of Conn2 or not?

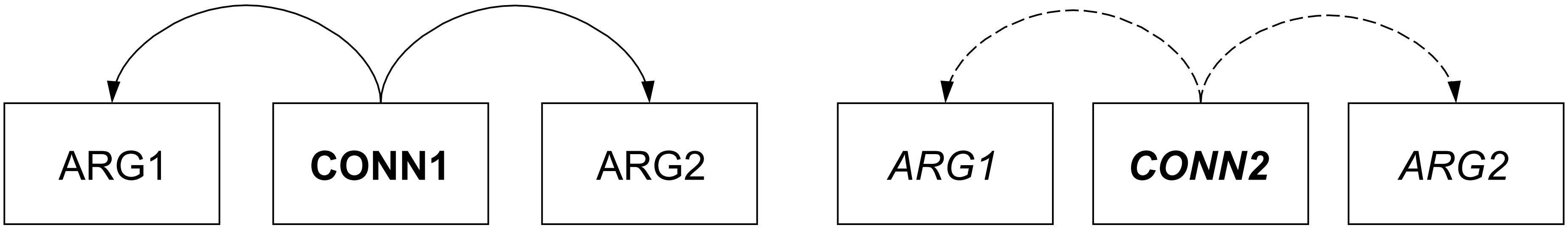
Spans of Consecutive Connectives

- **No common span among arguments to Conn1 and Conn2 (independent).**
- **Conn1 and its arguments are subsumed within an argument to Conn2, or vice versa (embedded).**
- **One or both arguments to Conn1 are shared with Conn2 (shared).**
- **One or both arguments to Conn1 overlap those of Conn2 (overlapping).**

Spans of Consecutive Connectives

- **Independent**
- **Embedded**
  - **Exhaustively Embedded**
  - **Properly Embedded**
- **Shared**
  - **Fully Shared**
  - **Partially Shared**
- **Overlapping**

Independent



Independent: Example

The securities-turnover tax has been long criticized by the West German financial community **BECAUSE** it tends to drive securities trading and other banking activities out of Frankfurt into rival financial centers, especially London, where trading isn't taxed. The tax has raised less than one billion marks annually in recent years, ***BUT*** the government has been reluctant to abolish the levy for budgetary concerns.

Independent: Example

ARG1

The securities-turnover tax has been long criticized by the West German financial community BECAUSE it tends to drive securities trading and other banking activities out of Frankfurt into rival financial centers, especially London, where trading isn't taxed. The tax has raised less than one billion marks annually in recent years, but the government has been reluctant to abolish the levy for budgetary concerns.

ARG2

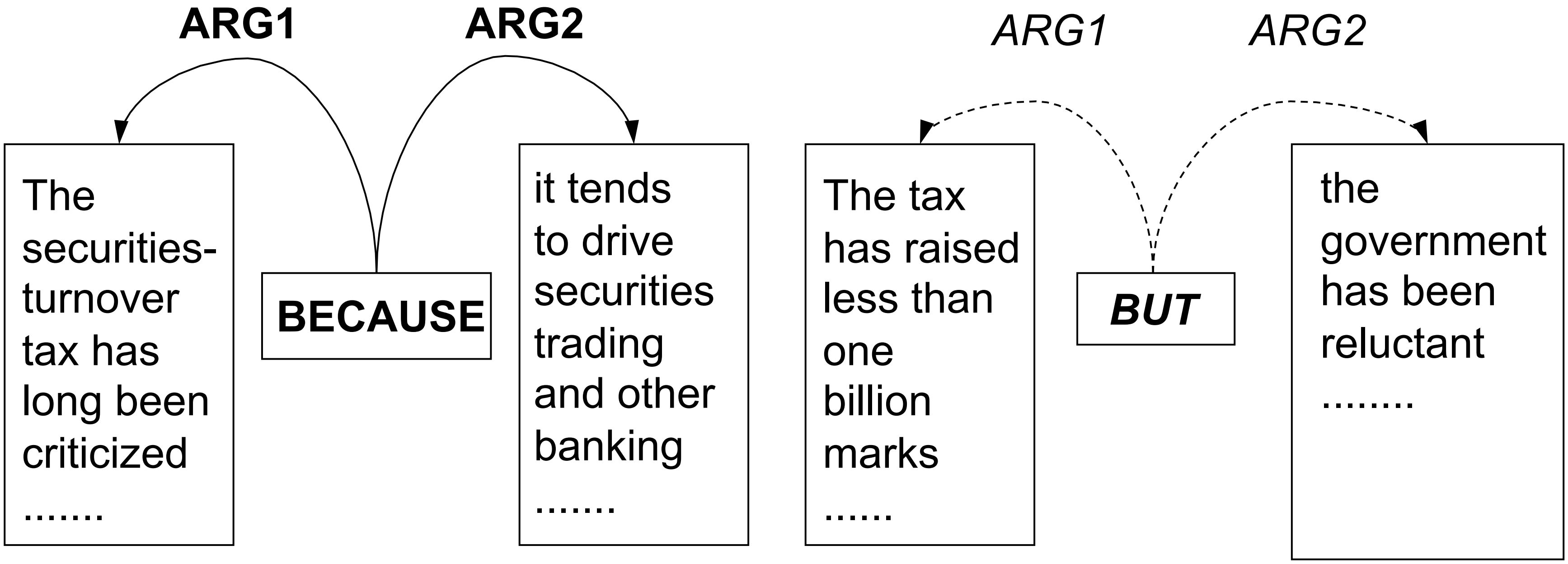
Independent: Example

ARG1

The securities-turnover tax has been long criticized by the West German financial community because it tends to drive securities trading and other banking activities out of Frankfurt into rival financial centers, especially London, where trading isn't taxed. **The tax has raised less than one billion marks annually in recent years**, ***BUT*** **the government has been reluctant to abolish the levy for budgetary concerns**.

ARG2

Independent: Example



ARG1

ARG2

The securities-turnover tax has long been criticized ....

BECAUSE

it tends to drive securities trading and other banking ....

ARG1

ARG2

The tax has raised less than one billion marks ......

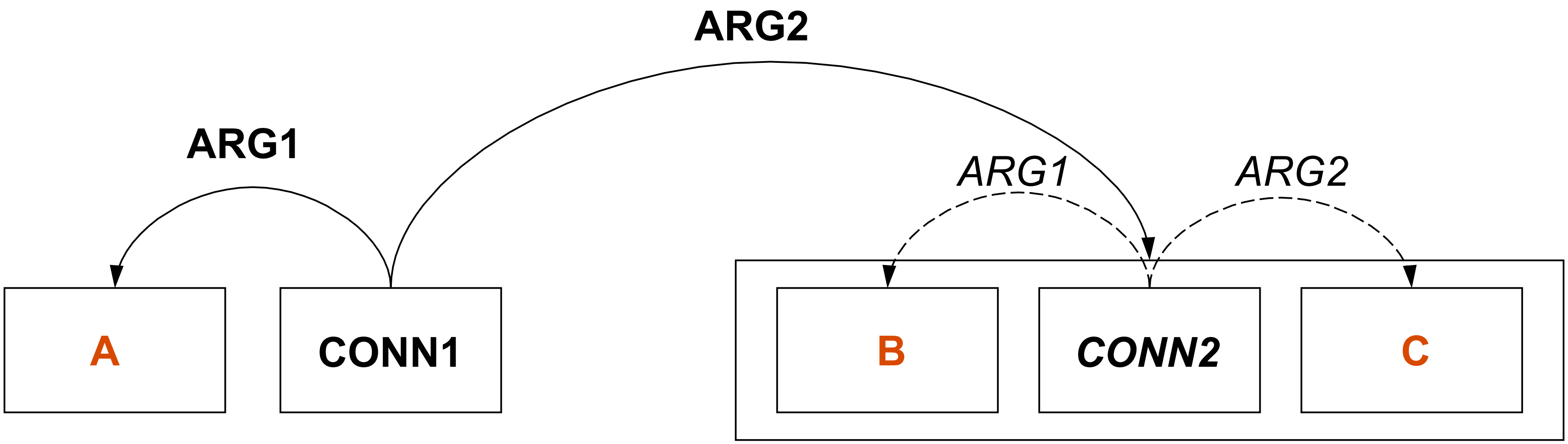
BUT

the government has been reluctant ....

Spans of Consecutive Connectives

- Independent
- **Embedded**
  - **Exhaustively Embedded**
  - Properly Embedded
- Shared
  - Fully Shared
  - Partially Shared
- Overlapping

Exhaustively Embedded



Exhaustively Embedded: Example

The drop in earnings had been anticipated by most Wall Street analysts, **BUT** the results were reported ***AFTER*** the market closed.

Exhaustively Embedded: Example

ARG1

The drop in earnings had been anticipated by most Wall Street analysts, BUT the results were reported after the market closed.

ARG2

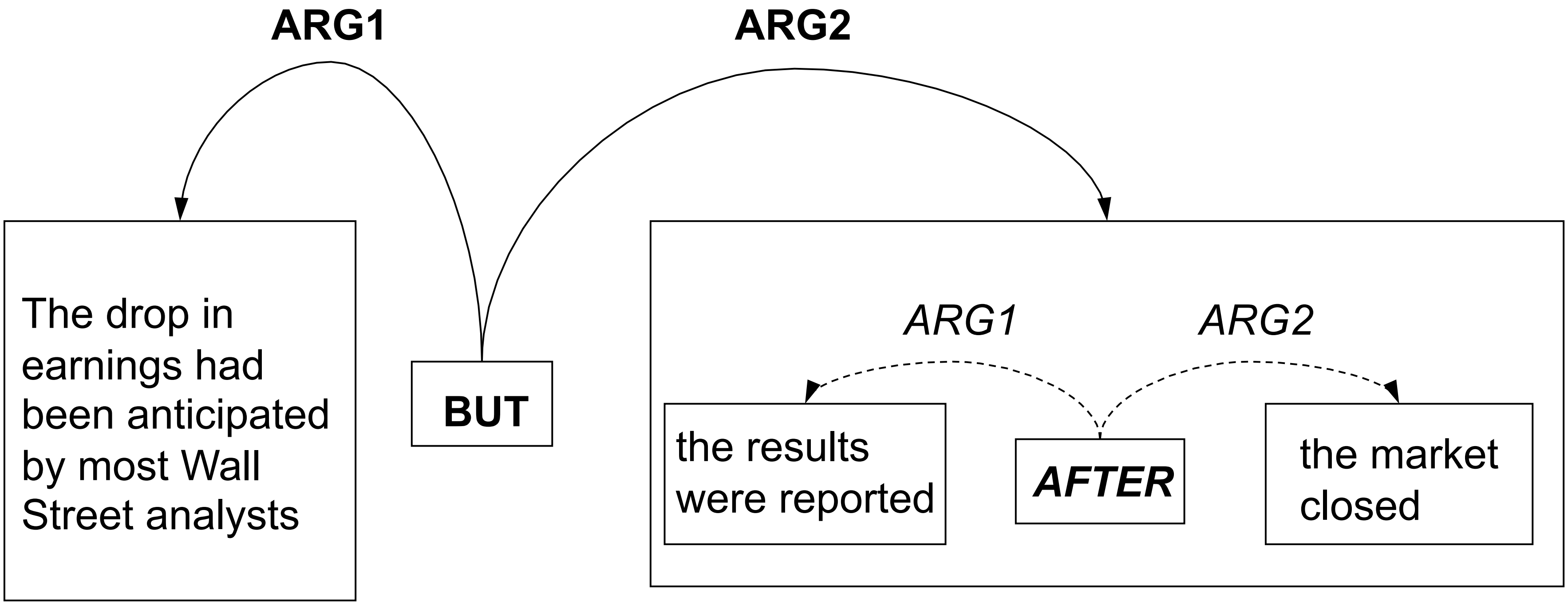
Exhaustively Embedded: Example

ARG1

The drop in earnings had been anticipated by most Wall Street analysts, but **the results were reported** ***AFTER*** **the market closed**.

ARG2

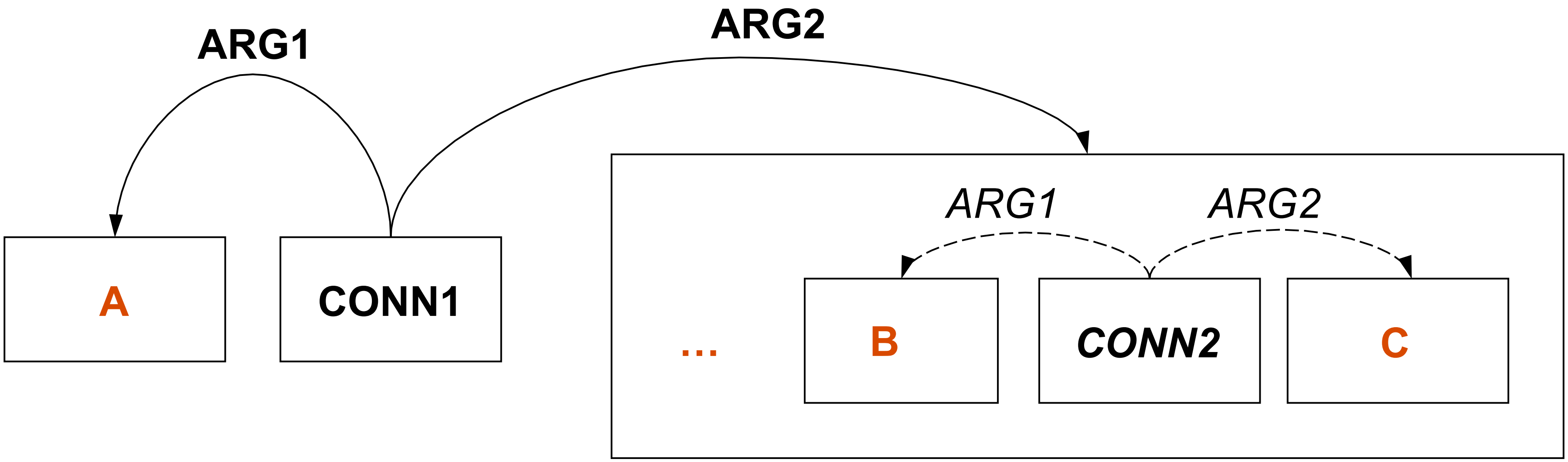
Exhaustively Embedded: Example


ARG1
ARG2
The drop in earnings had been anticipated by most Wall Street analysts
BUT
ARG1
ARG2
the results were reported
AFTER
the market closed

Spans of Consecutive Connectives

- **Independent**
- **Embedded**
  - **Exhaustively Embedded**
  - **Properly Embedded**
- **Shared**
  - **Fully Shared**
  - **Partially Shared**
- **Overlapping**

Properly Embedded



Properly Embedded: Example

The march got its major support from self-serving groups that know a good thing **WHEN** they see it, ***AND*** the crusade was based on greed or the profit motive.

Properly Embedded: Example

ARG1

The march got its major support from self-serving groups **that know a good thing** **WHEN** **they see it**, and the crusade was based on greed or the profit motive.

ARG2

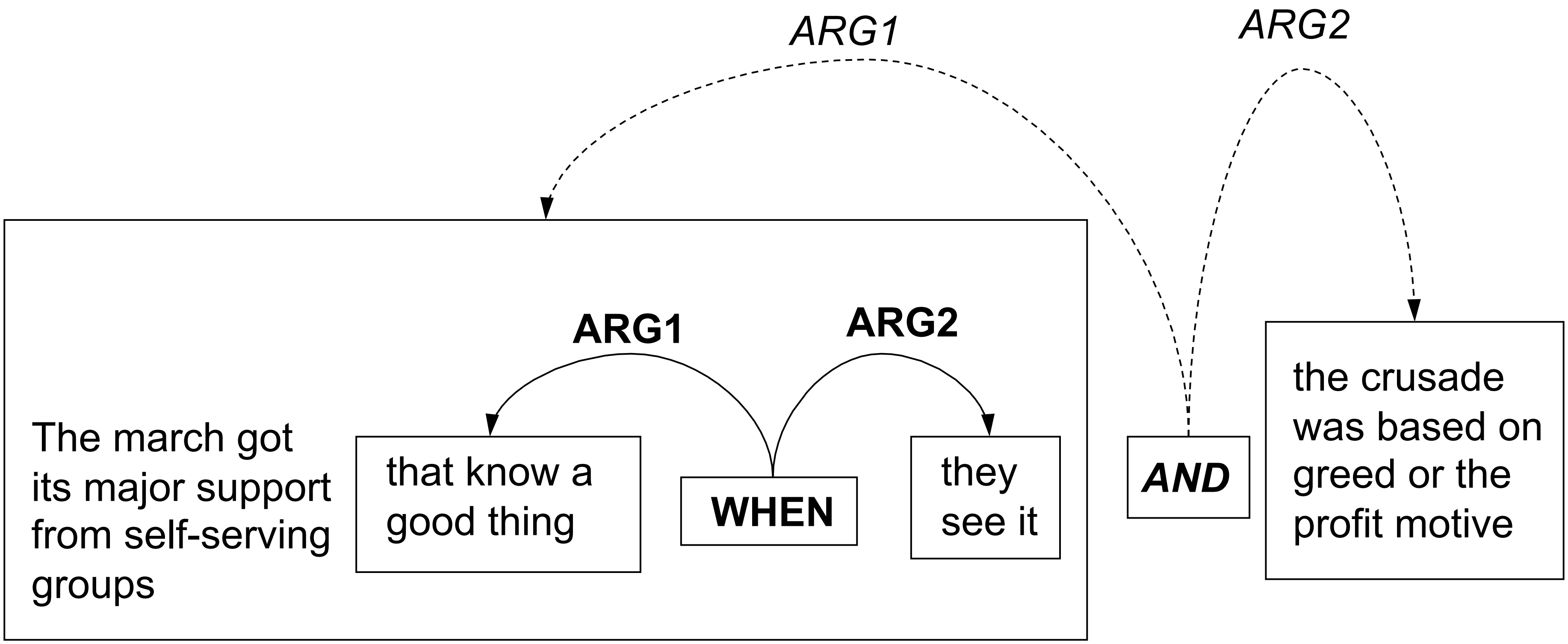
Properly Embedded: Example

ARG1

The march got its major support from self-serving groups that know a good thing when they see it, ***AND*** the crusade was based on greed or the profit motive.

ARG2

Properly Embedded: Example

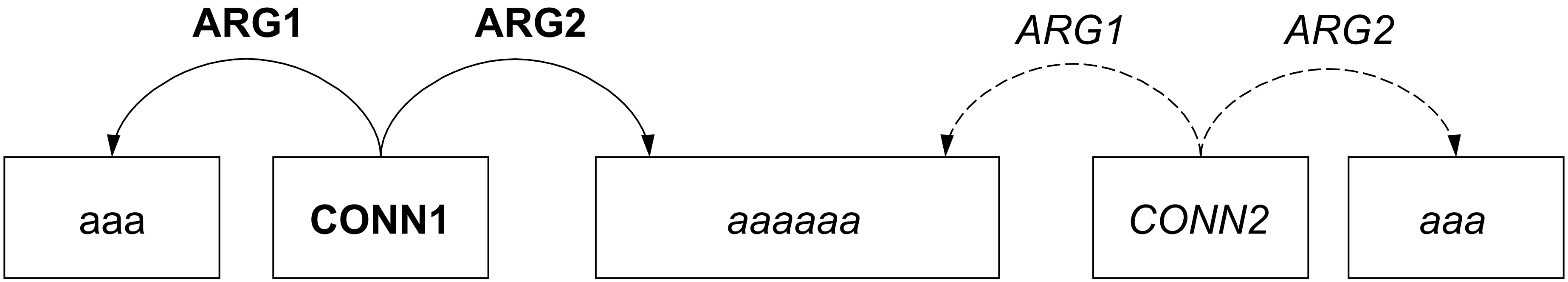


ARG1
ARG2
ARG1
ARG2
The march got its major support from self-serving groups
that know a good thing
WHEN
they see it
AND
the crusade was based on greed or the profit motive

Spans of Consecutive Connectives

- Independent
- Embedded
  - Exhaustively Embedded
  - Properly Embedded
- **Shared**
  - **Fully Shared**
  - Partially Shared
- Overlapping

Fully Shared Arg



Fully Shared Arg: Example

In times past, life-insurance companies targeted heads of household, meaning men, **BUT** ours is a two-income family and used to it. ***SO*** if anything happened to me, I'd want to leave behind enough so that my 33-year old husband would be able to pay off the mortgage and some other debts.

Fully Shared Arg: Example

ARG1

In times past, life-insurance companies targeted heads of household, meaning men, **BUT** **ours is a two-income family and used to it**. So if anything happened to me, I'd want to leave behind enough so that my 33-year old husband would be able to pay off the mortgage and some other debts.

ARG2

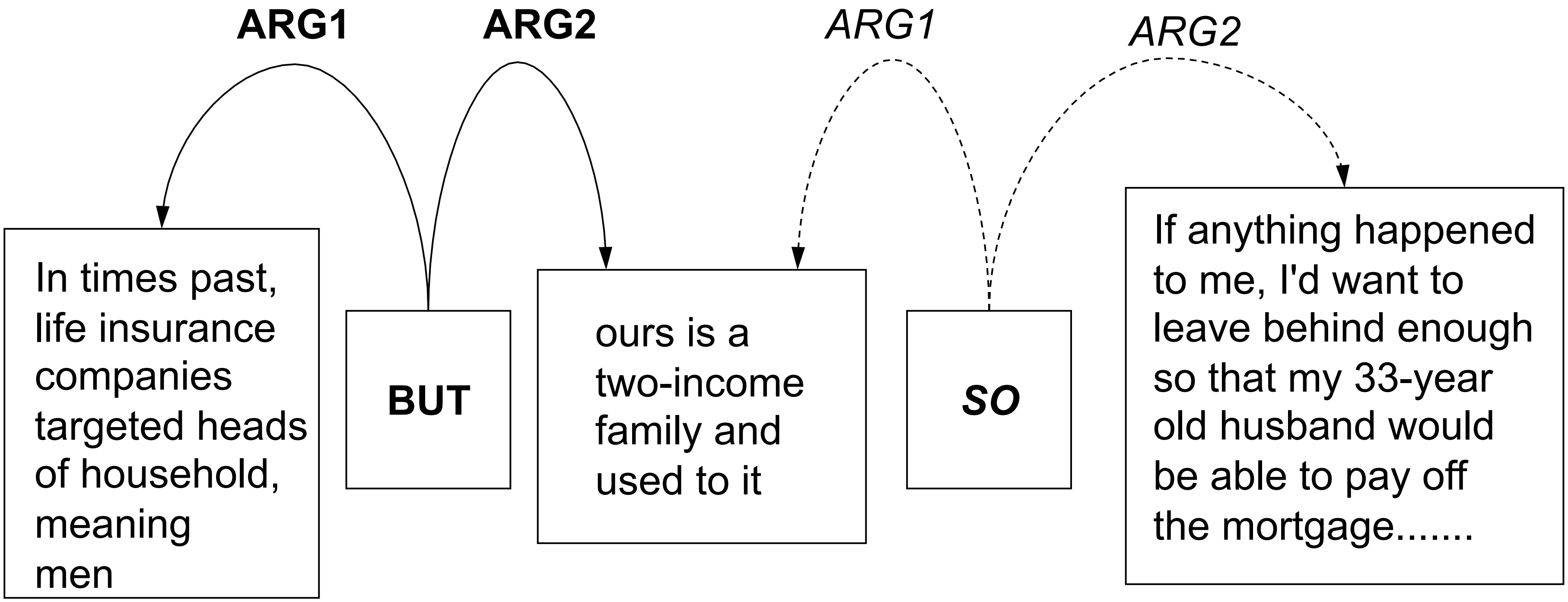
Fully Shared Arg: Example

In times past, life-insurance companies targeted heads of household, meaning men, but **ours is a two-income family and used to it**. ***SO*** **if anything happened to me, I'd want to leave behind enough so that my 33-year old husband would be able to pay off the mortgage and some other debts**.

ARG1

ARG2

Fully Shared Arg: Example



ARG1

ARG2

ARG1

ARG2

In times past, life insurance companies targeted heads of household, meaning men

BUT

ours is a two-income family and used to it

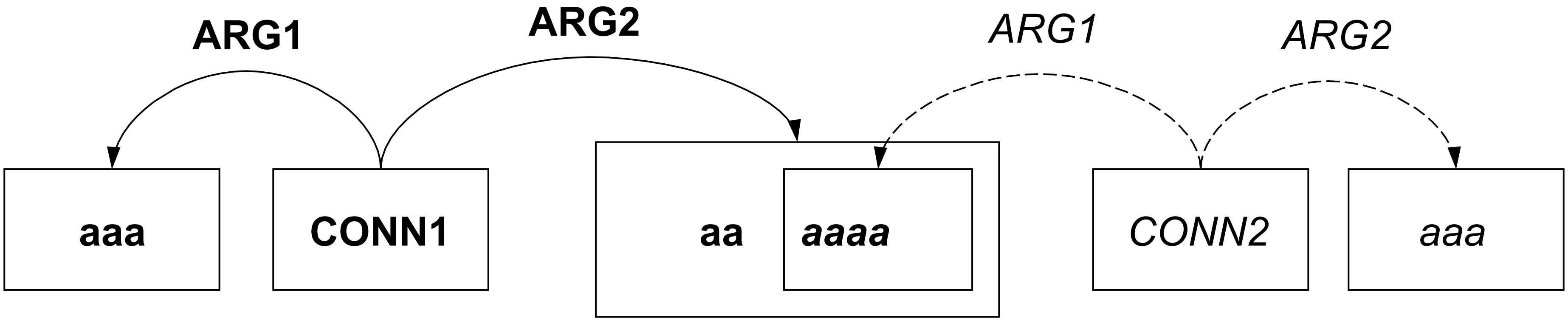
SO

If anything happened to me, I'd want to leave behind enough so that my 33-year old husband would be able to pay off the mortgage.......

Spans of Consecutive Connectives

- Independent
- Embedded
  - Exhaustively Embedded
  - Properly Embedded
- Shared
  - Fully Shared
  - Partially Shared
- Overlapping

Partially Shared Arg



Partially Shared Arg: Example

Japanese retail executives say the main reason they are reluctant to jump into the fray in the U.S. is that - unlike manufacturing - retailing is extremely sensitive to local cultures and life styles. **IMPLICIT=FOR EXAMPLE** The Japanese have watched the Europeans and Canadians stumble in the U.S. market, *AND* they fret that the business practices that have won them huge profits at home won't translate into success in the U.S.

Partially Shared Arg: Example

1st Discourse Relation

ARG1: that - unlike manufacturing - retailing is extremely sensitive to local cultures and life styles.

CONN: **FOR EXAMPLE**

ARG2: the Europeans and Canadians stumble in the U.S. market

Partially Shared Arg: Example

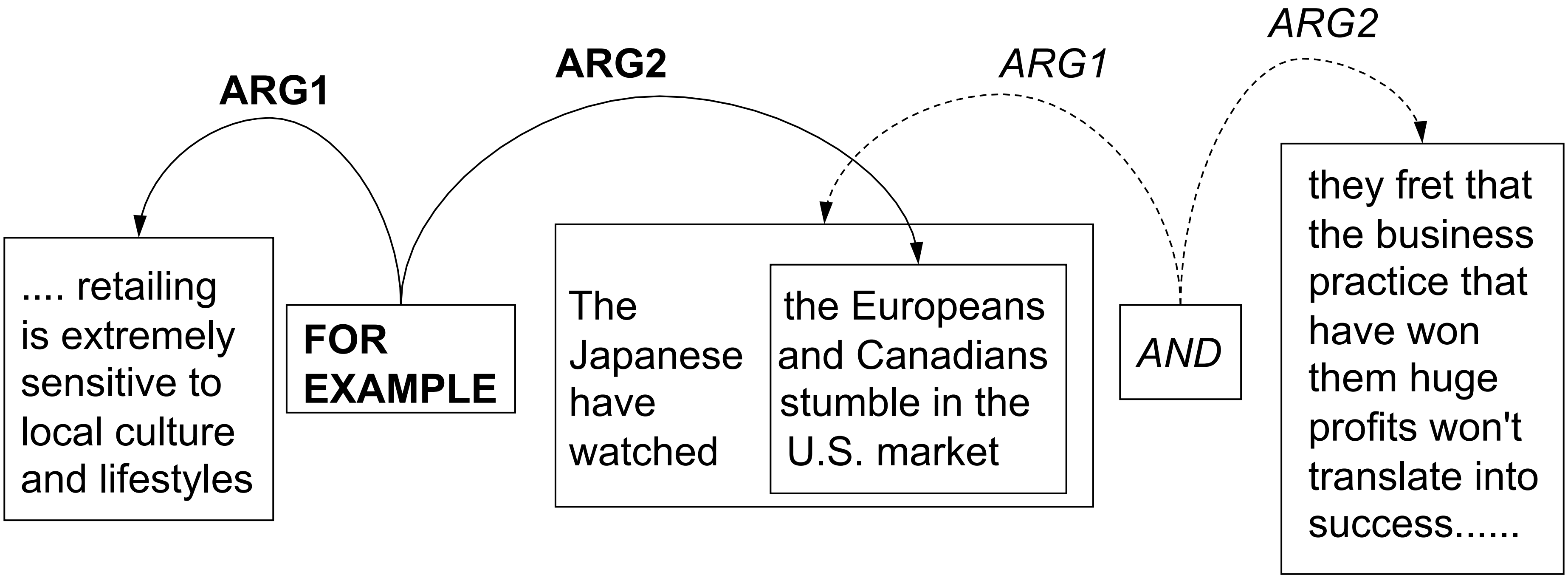
2nd Discourse Relation

ARG1: The Japanese have watched the Europeans and Canadians stumble in the U.S. market

CONN: *AND*

ARG2: they fret that the business practice that have won them huge profits at home won't translate into success in the U.S.

Partially Shared Arg: Example



.... retailing is extremely sensitive to local culture and lifestyles

FOR EXAMPLE

The Japanese have watched the Europeans and Canadians stumble in the U.S. market

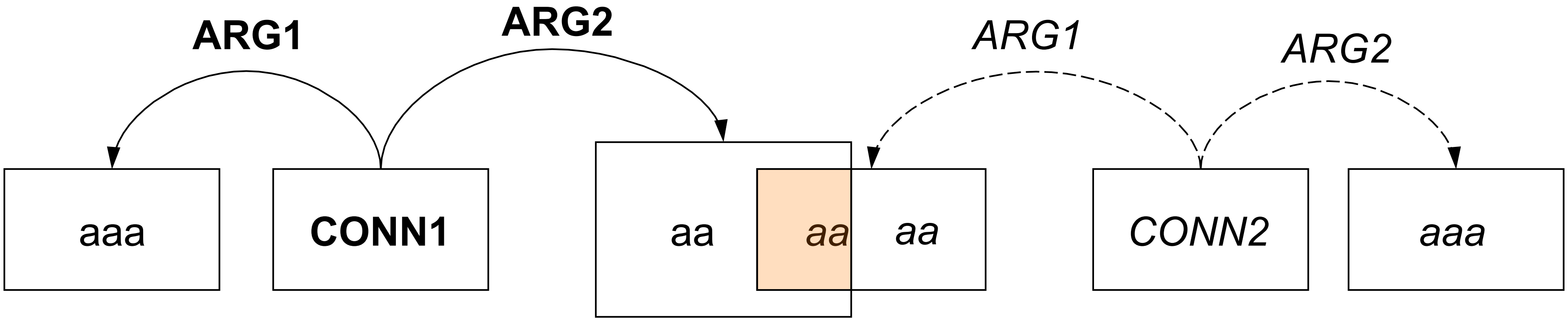
AND

they fret that the business practice that have won them huge profits won't translate into success......

Spans of Consecutive Connectives

- Independent
- Embedded
  - Exhaustively Embedded
  - Properly Embedded
- Shared
  - Fully Shared
  - Partially Shared
- Overlapping

Overlapping Args



Overlapping Args: Example

He (Mr. Meeks) said the evidence pointed to wrongdoing by Mr. Keating "and others," **ALTHOUGH** he didn't allege any specific violation. Richard Newsom, a California state official who last year examined Lincoln's parent, American Continental Corp, said he ***ALSO*** saw evidence that crimes had been committed.

Overlapping Args: Example

ARG1

He (Mr. Meeks) said the evidence pointed to wrongdoing by Mr. Keating "and others," **ALTHOUGH** **he didn't allege any specific violation**. Richard Newsom, a California state official who last year examined Lincoln's parent, American Continental Corp, said he also saw evidence that crimes had been committed.

ARG2

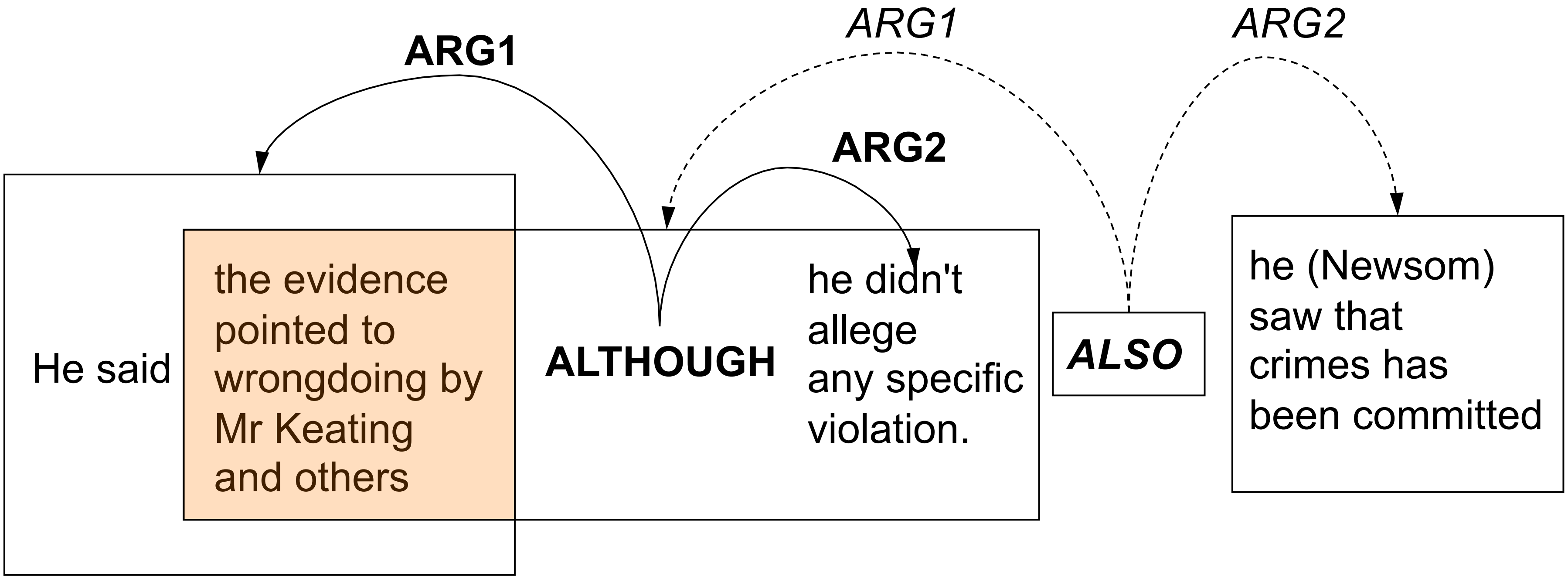
Overlapping Args: Example

ARG1

He (Mr. Meeks) said **the evidence pointed to wrongdoing by Mr. Keating "and others," although he didn't allege any specific violation**. Richard Newsom, a California state official who last year examined Lincoln's parent, American Continental Corp, said **he *ALSO* saw evidence that crimes had been committed**.

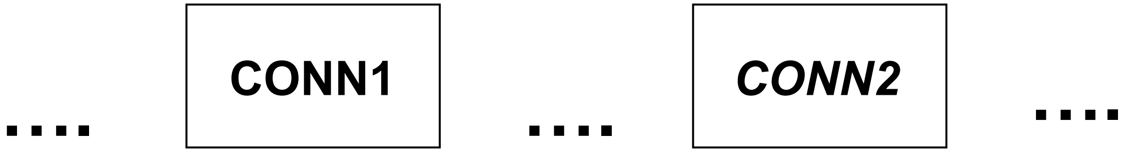
ARG2

Overlapping Args: Example



ARG1
ARG1
ARG2
ARG2
He said
the evidence pointed to wrongdoing by Mr Keating and others
ALTHOUGH
he didn't allege any specific violation.
ALSO
he (Newsom) saw that crimes has been committed

Pure Crossings

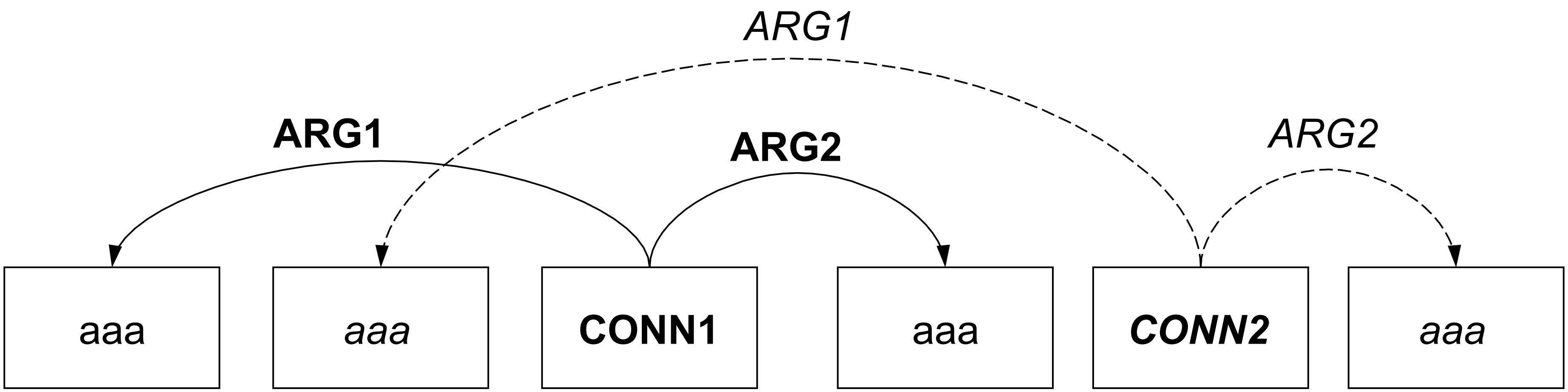


…. CONN1 …. *CONN2* ….

1. How do the text spans associated with Conn1 and its args relate to those of Conn2 and its args?

2. Do the pred-arg dependencies of Conn1 cross those of Conn2 or not?

Pure Crossing



Pure Crossing: Example

"I'm sympathetic with workers who feel under the gun," says Richard Barton of the Direct Marketing Association of America, which is lobbying strenuously against the Edwards beeper bill. "**BUT** the only way you can find out how your people are doing is by listening." The powerful group, which represents many of the nation's telemarketers, was instrumental in derailing the 1987 bill. Speigel ***ALSO*** opposes the beeper bill, saying the noise it requires would interfere with customer orders, causing irritation and even errors.

Pure Crossing: Example

ARG1

"**I'm sympathetic with workers who feel under the gun**," says Richard Barton of the Direct Marketing Association of America, which is lobbying strenuously against the Edwards beeper bill. "**BUT** **the only way you can find out how your people are doing is by listening**." The powerful group, which represents many of the nation's telemarketers, was instrumental in derailing the 1987 bill. Speigel also opposes the beeper bill, saying the noise it requires would interfere with customer orders, causing irritation and even errors.

ARG2

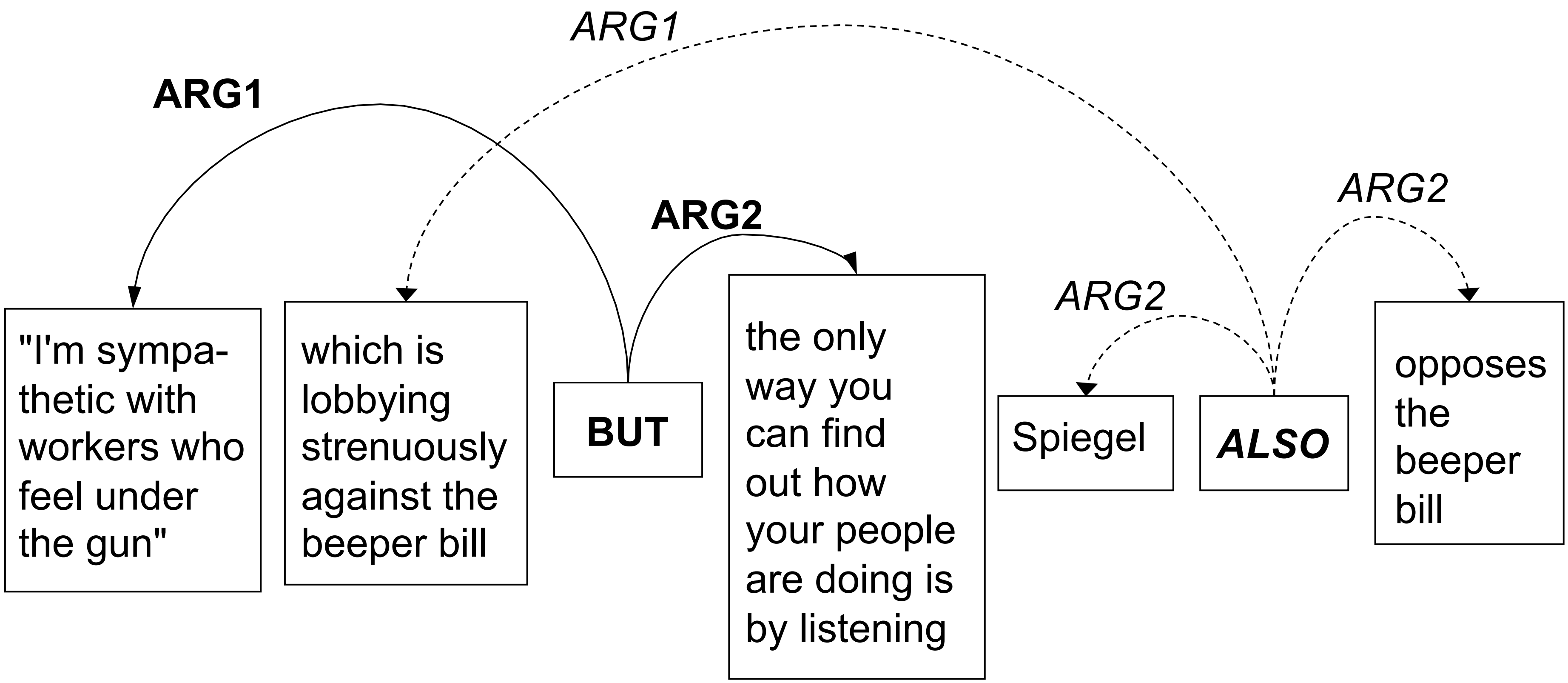
Pure Crossing: Example

ARG1

"I'm sympathetic with workers who feel under the gun," says Richard Barton of the Direct Marketing Association of America, **which is lobbying strenuously against the Edwards beeper bill**. "But the only way you can find out how your people are doing is by listening." The powerful group, which represents many of the nation's telemarketers, was instrumental in derailing the 1987 bill. **Spiegel** ***ALSO*** **opposes the beeper bill**, saying the noise it requires would interfere with customer orders, causing irritation and even errors.

ARG2

Pure Crossing: Example

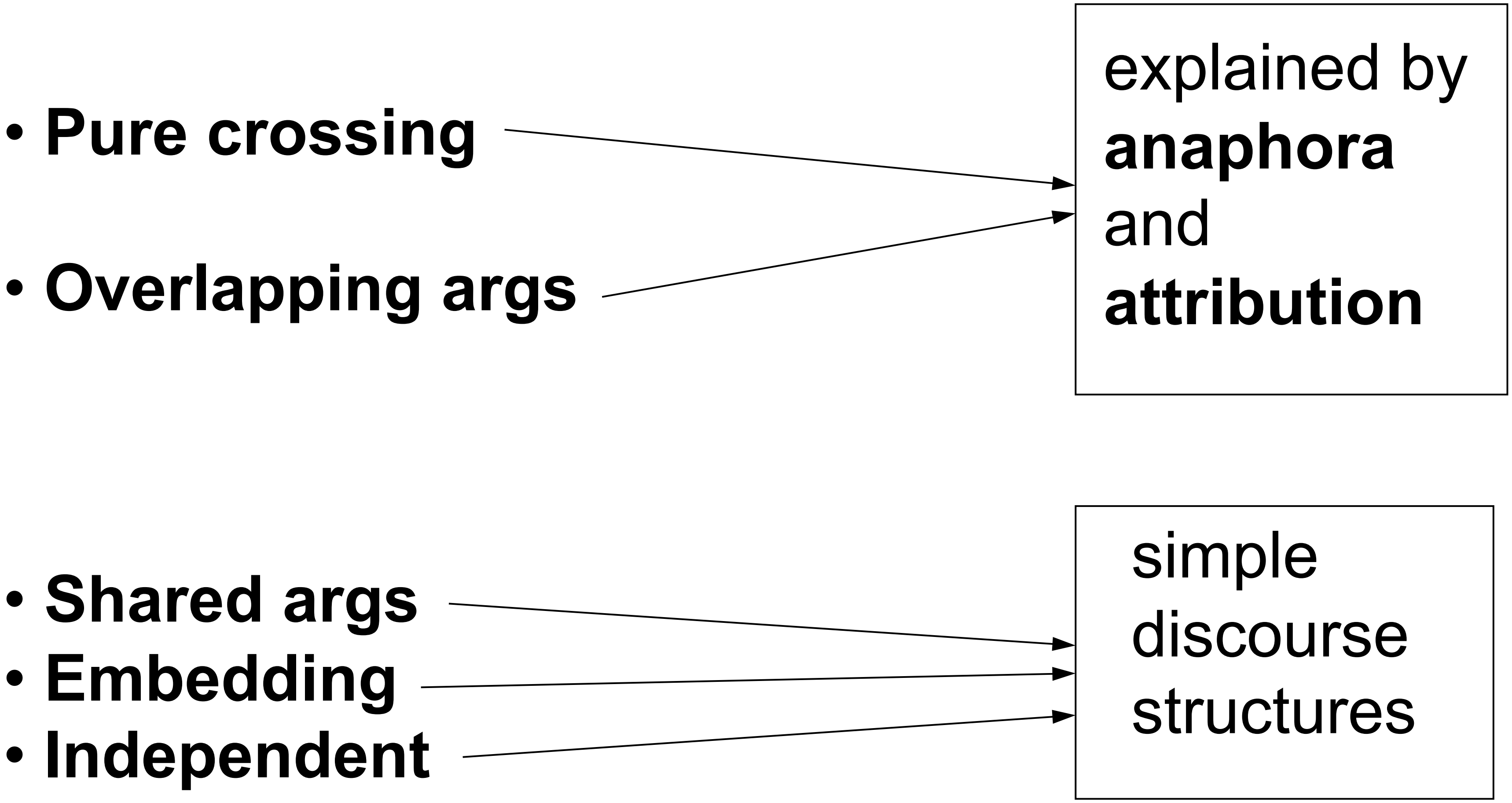

ARG1
ARG1
ARG2
ARG2
ARG2
"I'm sympathetic with workers who feel under the gun"
which is lobbying strenuously against the beeper bill
BUT
the only way you can find out how your people are doing is by listening
Spiegel
ALSO
opposes the beeper bill

Discussion

- Various grammar formalisms for syntax (e.g. LTAG) characterize certain crossing and nested (projective and non-projective) dependencies, leading to the so-called mildly context-sensitive languages.

- BUT in the PDTB corpus, we appear to see more complex discourse structures in English than we do in syntax. (Crossing dependencies, partially overlapping arguments, etc.) Is this a valid observation?

Explaining the Patterns of Consecutive Conns



- **Pure crossing**
- **Overlapping args**

explained by **anaphora** and **attribution**

- **Shared args**
- **Embedding**
- **Independent**

simple discourse structures

Discourse Anaphora and Pure Crossing

- All cases of pure crossing in the PDTB involve at least one **discourse adverbial**.

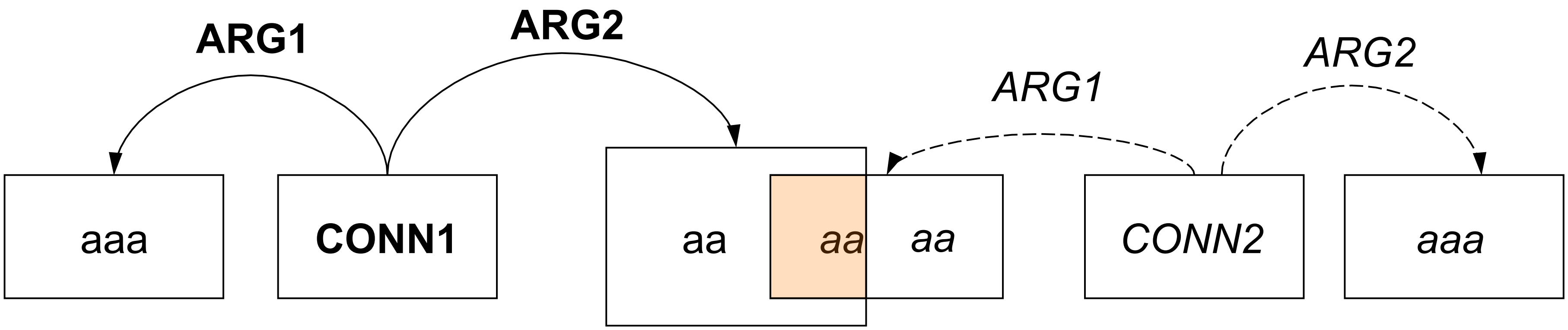
- With discourse adverbials, one argument is structural and the other is **anaphoric.**

- Anaphoric arguments are **NOT** specified structurally

  -- They are however annotated in PDTB

Overlapping Arguments: Explained by Attribution

The concept of **Attribution** explains the presence of Partially Overlapping Arguments in the PDTB.



Attribution

Attribution captures the relation of "ownership" between agents and Abstract Objects (arguments).

It is NOT a discourse relation (Mann & Thompson 1988). Attribution captures how discourse relations and their arguments can be attributed to different individuals:

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**.

RELATION and **Arg2** are attributed to the Writer.
Arg1 is attributed to another agent.

Attribution

Sometimes, the attribution predicates are simply part of the arguments:

ALTHOUGH *some lawyers reported* that prospective acquirers were scrambling to make filings before the fees take effect, *government officials said* they hadn't noticed any surge in filings.

Summary

- “Lexically” grounded annotation of discourse relations
- A brief description of the
  Penn Discourse Treebank (PDTB)
  PDTB 2.0 to be available around November 2007
- Annotations of discourse connectives (explicit and implicit), attributions, and senses of connectives
- Moving towards discourse meaning
- Annotations specify structures over parts of the discourse and not necessarily all the discourse
  -- compare with syntactic annotation
- *Complexity of dependencies at the discourse level may be no more than that in PDTB, even for languages for which the complexity at the syntactic level is greater than the syntactic complexity for English*

Do we want a single tree over a sentence?

- **There are many constructions in language that suggest that the single tree hypothesis may be wrong**
  - -- **Parentheticals, supplements, sentential relatives, among others are problematic for the single tree hypothesis**

Mary, John thinks, will win the election

(John thinks is attached to the S node medially but it has scope over Mary will win the election)

John heard that
Mary finally finished her dissertation,
which no one ever expected her to do so

((1) John heard that and (2) which no one ever expected her to do both have scope over (3) Mary finally finished her dissertation. Both (1) and (2) are attached to the root node S but neither (1) nor (2) have scope over the other)

Alternative Lexicalization (AltLex)

A discourse relation is inferred between two sentences which do not contain an Explicit connective, but insertion of an Implicit connective leads to redundancy. This is because the relation is **alternatively lexicalized** by some non-connective expression:

- ***Under a post-1987 crash reform, the Chicago Mercantile Exchange wouldn't permit the December S&P futures to fall further than 12 points for a half hour***. **AltLex = (consequence)** **That caused a brief period of panic seeling of stocks on the Big Board**.

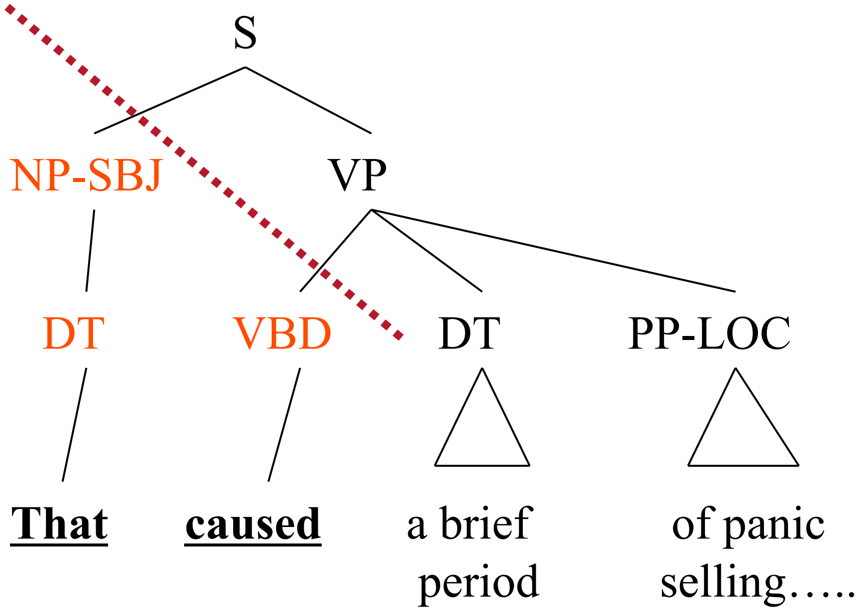
Discourse Connectives and Syntactic Constituency

- Most explicit connectives correspond to syntactic constituencies. E.g. (“because” IN, “but” CC, “as a result” PP, etc.)

- Some small exceptions with parallel connectives, as we have seen.

AltLex expressions often do not correspond to syntactic constituencies.

Under a post-1987 crash reform, the Chicago Mercantile Exchange wouldn't permit the December S&P futures to fall further than 12 points for a half hour. **AltLex = (consequence)** **That caused a brief period of panic selling of stocks on the Big Board**.



For a list of AltLex expressions annotated in the PDTB:

http://www.seas.upenn.edu/~pdtb/altlex-strings.txt

Or search using PDTB Browser:

http://www.seas.upenn.edu/~pdtb/PDTBAPI/pdtbbrowser.jnlp

Explicit connectives are the lexical items that trigger discourse relations.

- Subordinating conjunctions (e.g., *when*, *because*, *although,* etc.)
  - ➢ ***The federal government suspended sales of U.S. savings bonds*** **because** **Congress hasn't lifted the ceiling on government debt**.
- Coordinating conjunctions (e.g., *and*, *or*, *so*, *nor*, etc.)
  - ➢ ***The subject will be written into the plots of prime-time shows***, **and** **viewers will be given a 900 number to call**.
- Discourse adverbials (e.g., *then*, *however*, *as a result*, etc.)
  - ➢ ***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**, **industry operated out of small, expensive, highly inefficient industrial units**.

- Only 2 AO arguments, labeled ***Arg1*** and **Arg2**
- **Arg2**: clause with which connective is syntactically associated
- ***Arg1***: the other argument

How is Chinese the same (as English)?

- Explicit and implicit relations
- Explicit connectives (Xue 2005):
  - Subordinating conjunctions
  - Coordinating conjunctions
  - Discourse adverbials

Chinese discourse connectives: classification

- Subordinate conjunctions
- Coordinate conjunctions
- Adverbial connectives
- Implicit connectives
- Localizers (?)

Subordinate conjunctions

- **如果/if** 改革/reform 措施/measure 不/not 得力/effective (**那么/then**) 投资者/investor **就/then** 有/have 可能/possibility 把/BA 注意力/attention 转向/turn to 新兴/emerging 市场/market。

  "If the reform measures are not effective, confidence crisis still exists, then investors are likely to turn their attention to other emerging markets."

Coordinate conjunctions

- 现代/modern 父母/parent 难/difficult 为/to 的/DE 地方/area 是/be **既/not only** 无法/no way 排除/eliminate 血液/blood 中/in 传统/traditional 的/DE 观念/values **又/but also** 要/need 面对/face 新/new 的/DE 价值/values 。

  "The difficulty of being modern parents lies in the fact they can not get rid of the traditional values flowing in their blood, and they also need to face new values."

Adverbial connectives

- 克林顿/Clinton 政府/Admininstration 已经/already 表示/indicate 要/will 延长/extend 中国/China 的/DE 最惠国/MFN 待遇/status, **因此/theorefore** 这/this 次/CL 游说/lobby 的/de 对象/target 是/be 那些/those 较/relatively 保守/conservative 的/DE 议员/congressmen。

  “The Clinton Administration has already indicated that it will extend China's MFN status, therefore, the focus of the lobby this time is on those relatively conservative congressmen.”

Implicit connectives

- 出口/export 比/compared with 去年/last year 下降/decrease 百分之一点三/1.3%, (**而/while**)进口/import 比/compared with 去年/last year 增长/increase 百分之三十四点一/34.1%。

  ”Export decreased 1.3 percent over the same period last year while import incresed 34.1 percent."

- Where possible, substitute an explicit connective

It can get a little tricky…

- 台商/Taiwan businessmen 子弟/children 学校/school **(虽然/although)** 已经/already 奠基/lay foundation，**但/{but, however}** 经费/funding 不足/insufficient，师资/faculty 未定/undecided。

  “The foundation of the school for Taiwan businessmen has been laid, but the funding is insufficient and its faculty hasn’t been decided.”

  – Subordinate conjunction? Coordinate conjunction? Adverbial connective?

Chinese discourse connectives: sense disambiguation

- 而 (“er”)
  - While
  - And
  - But
  - Instead
  - In addition
  - Other non-discourse connective senses

Chinese discourse connectives: Sense disambiguation

一九九七年发达国家的经济形势是美国经济增长强劲而日本经济疲软。

"The economic situation in developed countries in 1997 is that the U.S. (economy) grows strongly while the Japanese economy is weak."

水东开发区是适应乙烯工程需要而建立的后续加工基地。

"Shuidong Development Zone is a downstream processing base established to meet the need of the ethylene project."

能生产中国不能生产而又十分需要的药品的企业

"Enterprises that can produce drugs that China badly needs but cannot produce"

国际社会的参与对浦东的开发开放起了积极而关键的作用。

"The participation of the international community played a positive and key role in Huichun's development and opening up to the outside."

这不是历史的巧合，而是历史的积累转接。

"This certainly is not historical coincidence. Instead it is historical accumulation and transition."

Chinese discourse connectives: variation

Gloss	Part 1	Part 2
although	虽然, 虽说, 虽	但是,但,还是,可是,却,然而,不过
because	因为,因,由于	所以,故,而
if	如果,若,假如	那么,就
even if	即使	也,仍然,仍,还是,依然
as long as	只要	就,即
only if	只有	才
therefore	于是,因此	
for example	如例,如	

(lots of) Parallel connectives in Chinese

伦敦 股市 因 适逢 银行节，
London stock market because coincide Bank Holiday ,
故 没有 开市。
therefore NEG open market

“London Stock Market did not open because it was Bank Holiday.”

虽然 他们 不 离 土 、不 离 乡， 但 严格
Although they not leave land , not leave home village , but strictly
来 讲 已 不再 是 传统 意义 上 的 农民。
PART speak already no longer be tradition sense PREP DE peasant

“Although they do not leave land or their home village, strictly speaking, they are no longer peasants in the traditional sense.”

How is Chinese different?

- Complex ideas in one sentence, intra-sentential discourse relations often delimited by comma without an explicit connective
- Significantly more implicit relations than in English: 82% implicit in Chinese vs. 54.5% implicit in PDTB 2.0
- Discourse connectives are often optional

Adaptations

- Complex ideas in one sentence, intra-sentential discourse relations often delimited by comma without an explicit connective
  - Use commas as well as periods as indicators of discourse relations
- Significantly more implicit relations than in English: 82% implicit in Chinese vs. 54.5% implicit in PDTB 2.0
  - Annotate explicit and implicit discourse connectives in one unified process
- Discourse connectives are often optional
  - Define argument labels semantically

A Chinese Sentence

据悉， 东莞 海关 共 接受 企业
According to reports, Dongguan Customs in total accept company
合同 备案 八千四百多份， 比 试点 前 略 有
contract record 8400 plus CL, compare pilot before slight EXIST
上升， 企业 反应 良好， 普遍
increase , company respond/response good/well , generally
表示 接受 。
acknowledge accept/acceptance .

“According to reports, Dongguan District Customs accepted more than 8400 records of company contracts, a slight increase from before the pilot. Companies responded well, generally acknowledging acceptance.”

Commas as indicators for discourse units

据悉， [$_{AO1}$ 东莞 海关 共 接受 企业

According to reports, Dongguan Customs in total accept company

合同 备案 八千四百多份， 比 试点 前 略 有

contract record 8400 plus CL, compare pilot before slight EXIST

上升]， [$_{AO2}$ 企业 反应 良好， 普遍

increase , company respond/response good/well , generally

表示 接受] 。

acknowledge accept/acceptance .

“According to reports, [$_{AO1}$Dongguan District Customs accepted more than 8400 records of company contracts, a slight increase from before the pilot]. [$_{AO2}$Companies responded well, generally acknowledging acceptance].”

Or should it be…

据悉， [$_{AO1}$ 东莞 海关 共 接受 企业
According to reports, Dongguan Customs in total accept company
合同 备案 八千四百多 份]，[$_{AO2}$ 比 试点 前 略 有
contract record 8400 plus CL, compare pilot before slight EXIST
上升]，[$_{AO3}$ 企业 反应 良好]，[$_{AO4}$普遍
increase , company respond well , generally
表示 接受]。
acknowledge acceptance .

“According to reports, [$_{AO1}$Dongguan District Customs accepted more than 8400 records of company contracts], [$_{AO2}$ a slight increase from before the pilot]. [$_{AO3}$Companies responded well], [$_{AO4}$ generally acknowledging acceptance].”

Discourse relations

据悉， [$_{AO1}$ 东莞 海关 共 接受 企业
According to reports, Dongguan Customs in total accept company
合同 备案 八千四百多 份]，[$_{AO2}$ 比 试点 前 略 有
contract record 8400 plus CL, compare pilot before slight EXIST
上升]，[$_{AO3}$ 企业 反应 良好]，[$_{AO4}$ 普遍
increase , company respond well , generally
表示 接受] 。
acknowledge acceptance .

“According to reports, [$_{AO1}$ Dongguan District Customs accepted more than 8400 records of company contracts]. [$_{AO2}$ a slight increase from before the pilot]. [$_{AO3}$ Companies responded well], [$_{AO4}$ generally acknowledging acceptance].”

AO1 ~ AO2 ?
AO3 ~ AO4 ?
AO1 ~ AO3 ?

Explicit and implicit connectives unified

- Use punctuations as potential indicators of discourse relation
  - 82% implicit, 18% explicit
  - use of a discourse connective is almost always accompanied by a punctuation or two
- Mark explicit connectives as an attribute of the discourse relation

Arg1/2 Defined Semantically

- Why?
  - 82% implicit: distinction less meaningful
  - Discourse connectives often optional
- How?
  - Use the sense hierarchy already developed for English
  - Example: CONTINGENCY: Cause
    - “reason”: for cases like *because, since* etc.
    - “result”: for cases like *so, as a result* etc.
    - “reason”➔Arg1, clause bound to 因 (“*because”)* etc.
    - “result” ➔Arg2, clause bound to 故 (“*therefore”)* etc.

Arg1/2 Defined Semantically (Scheme for Chinese)

- Why?
  - 82% implicit: distinction meaningless
  - Parallel connectives prevalent: duplication
  - In Chinese, the old definition of Arg1/2 is meaningless in most cases; in the remaining cases, it often leads to duplication
- How?
  - Sense hierarchy
  - The “reason”/“result” distinction as intrinsic property of the type “CONTINGENCY:Cause”, not as subtypes
    - “reason”➜Arg1, clause bound to 因 (“*because”)* etc.
    - “result” ➜Arg2, clause bound to 故 (“*therefore”)* etc.

References

- Eleni Miltsakaki, Rashmi Prasad, Aravind Joshi, and Bonnie Webber. 2004. Annotating discourse connectives and their arguments. In *Proceedings of the HLT/NAACL Workshop on Frontiers in Corpus Annotation*, pages 9–16, Boston, MA, May.
- Rashmi Prasad, Eleni Miltsakaki, Nikhil Dinesh, Alan Lee, Aravind Joshi, Livio Robaldo, and Bonnie Webber, 2007. *The Penn Discourse Treebank 2.0 Annotation Manual*. The PDTB Research Group, December.
- Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.
- Nianwen Xue, Fei Xia, Fu dong Chiou, and Martha Palmer. 2005. The Penn Chinese TreeBank: Phrase Structure Annotation of a Large Corpus. *Natural Language Engineering*, 11(2):207–238.
- Nianwen Xue. 2005. Annotating the Discourse Connectives in the Chinese Treebank. In *Proceedings of the ACL Workshop on Frontiers in Corpus Annotation*, Ann Arbor, Michigan.