Dependencies at the Sentence Level and at the Discourse Level

Aravind K. Joshi

Department of Computer and Information Science

and

Institute for Research in Cognitive Science

University of Pennsylvania

Prague

CLARA WORKSHOP, Prague

December 15-17 2010

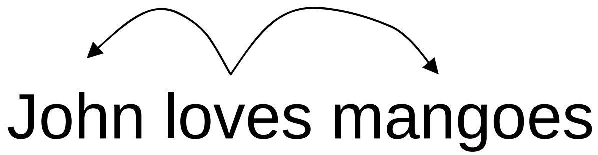
Outline

- Types of Dependencies
  -- word-word, word-phrase (text span), phrase-word, and phrase-phrase
- Dependencies at the Sentence Level
- Dependencies at the Discourse Level
  as illustrated by some examples from PDTB
- Comparison of Dependencies at the Sentence Level
  and at the Discourse Level
- What can we learn from the dependencies at the discourse level that may make us change our representations of structure at the sentence level
- Implications for Semantics

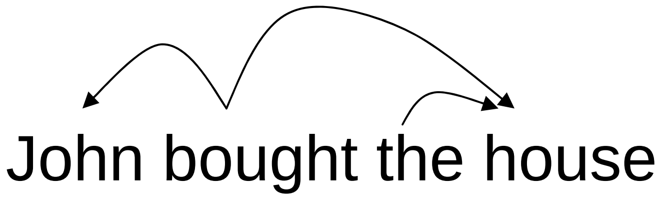
Summary

Types of Dependencies

- **Word to Word**



John loves mangoes

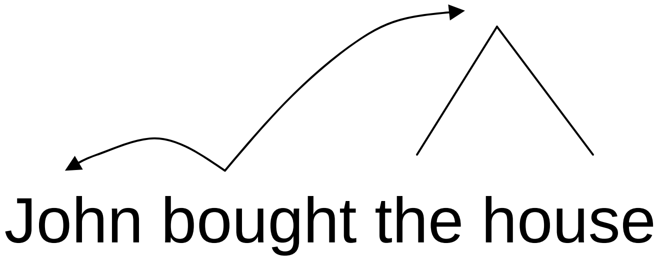


John bought the house

Predicate argument relation?

Types of Dependencies

Word to Phrase

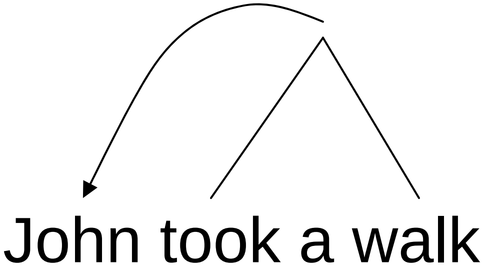


John bought the house

Predicate argument relation?

Types of Dependencies

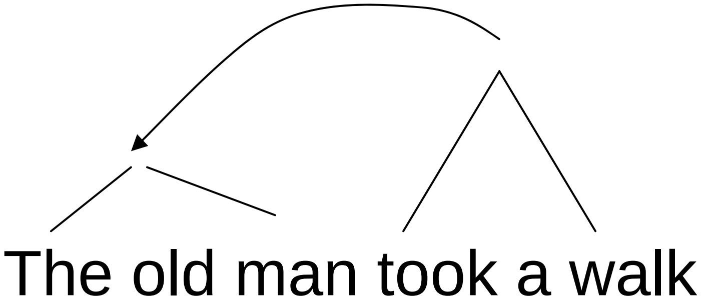
Phrase to Word



John took a walk

Types of Dependencies

Phrase to Phrase



The old man took a walk

Types of Dependencies

How much of the phrase to be included in the argument?

By convention (?) we take the maximal phrase.

John bought [the house next door which was on sale for over a year]

the house
the house next door
the house next door which was on sale for over a year

What about the minimal phrase that is sufficient to identify
the referent in the context (discourse context, for example)?

Comparing and Contrasting Dependencies at the Sentence Level and at the Discourse Level

- A very fast description of PDTB
- Possible Implications for Annotations at the Sentence Level

Penn Discourse Treebank (PDTB)

- **Wall Street Journal (same as the Pen Treebank (PTB) corpus): ~1M words**
  - **Annotations record**
- **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 (May 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**.

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

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**

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=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.**

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:

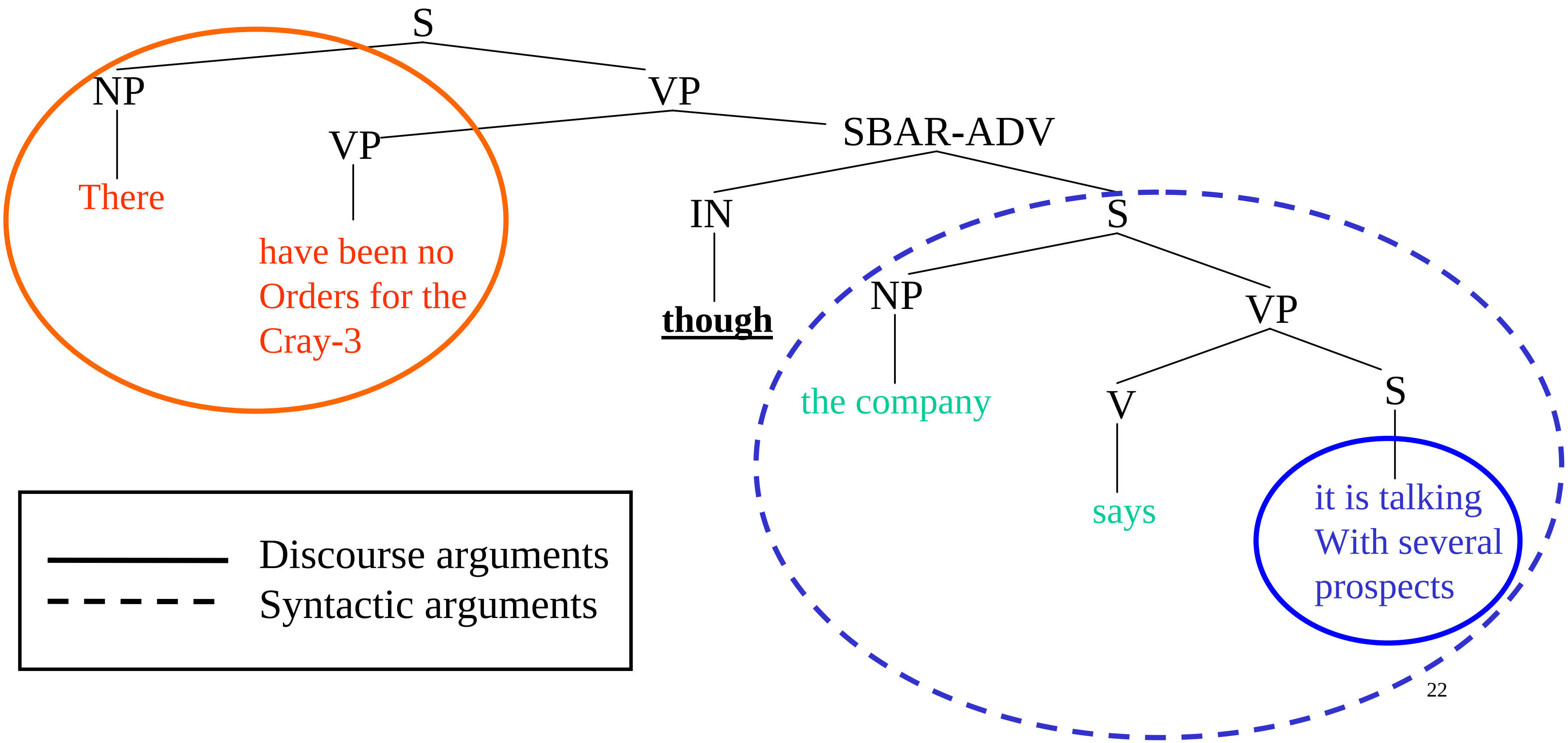
(1) How discourse relations and their arguments can be *attributed to different individuals*:

- 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*.

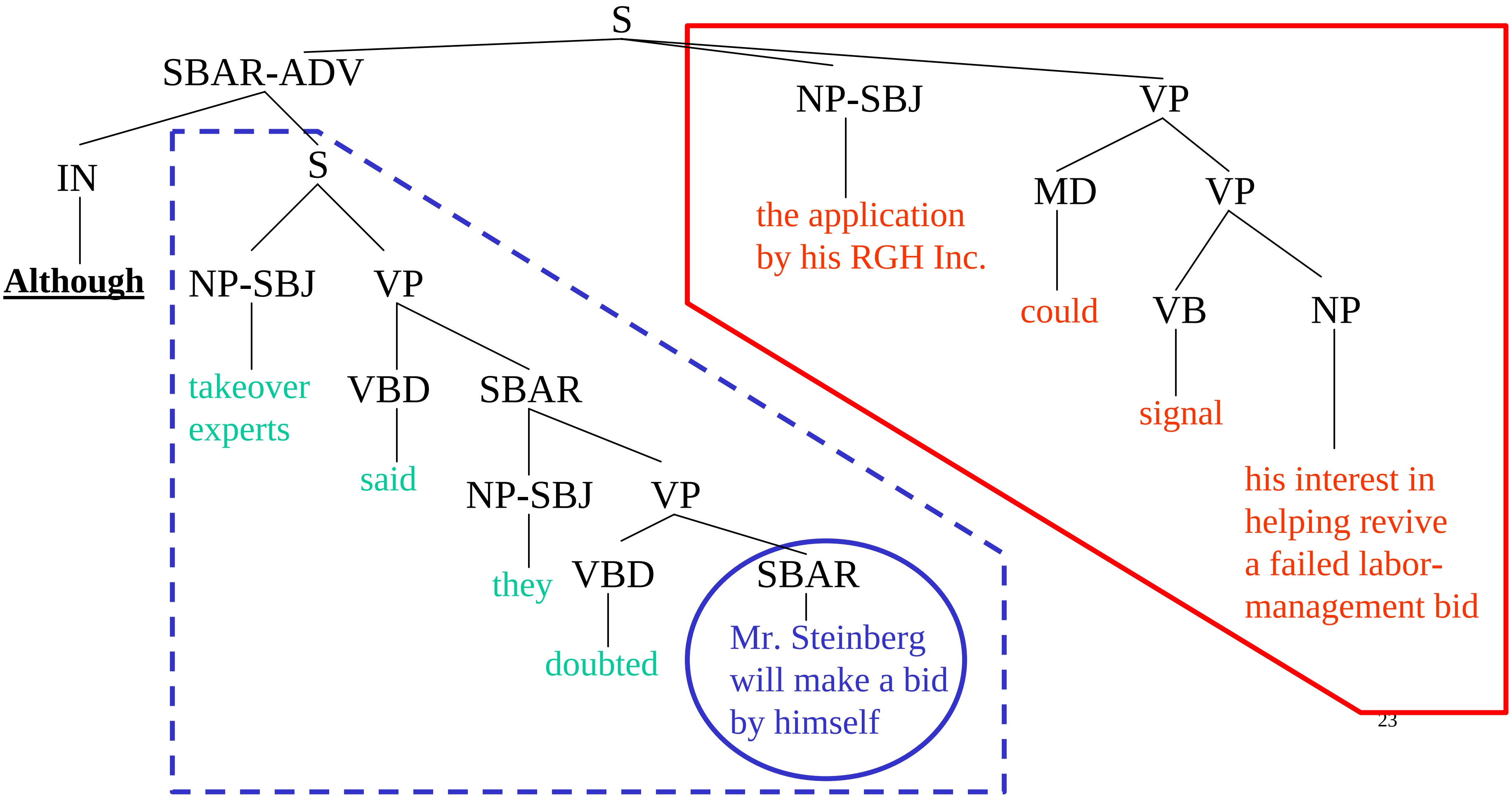
⇒ Relation and Arg2 are attributed to the Writer.

⇒ *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 **because** 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.

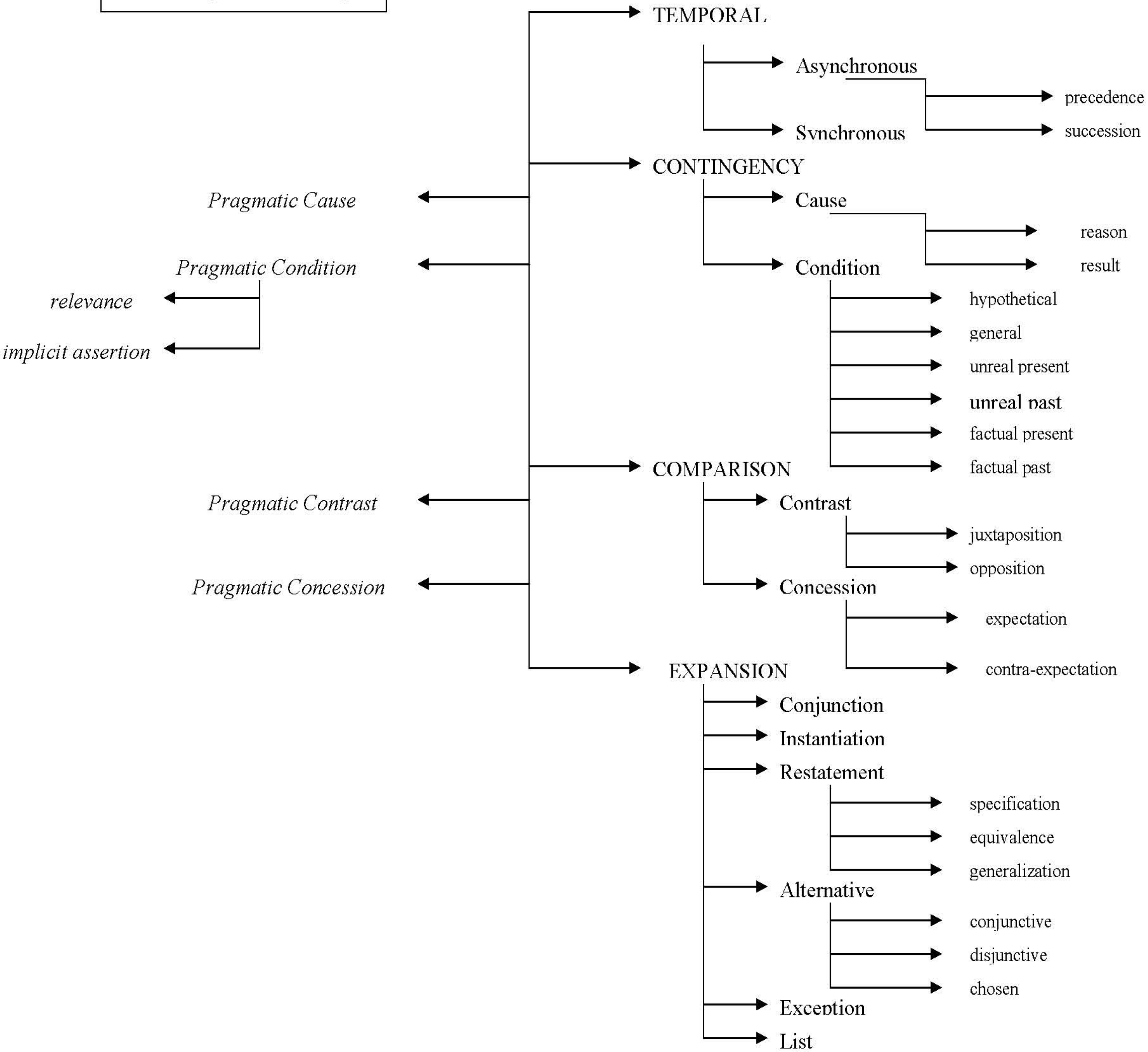
✓ Discourse semantics: causal relation between “investors being nervous” and “problems with the company’s ability to meet debt payments”

✖ 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, **while** opponents argued that the increase will still hurt small business and cost many thousands of jobs.

Hierarchy of sense tags



- TEMPORAL
  - Asynchronous
    - precedence
    - succession
  - Synchronous
- CONTINGENCY
  - Cause
    - reason
    - result
  - Condition
    - hypothetical
    - general
    - unreal present
    - unreal past
    - factual present
    - factual past
  - *Pragmatic Cause*
  - *Pragmatic Condition*
    - *relevance*
    - *implicit assertion*
- COMPARISON
  - Contrast
    - juxtaposition
    - opposition
  - Concession
    - expectation
    - contra-expectation
  - *Pragmatic Contrast*
  - *Pragmatic Concession*
- EXPANSION
  - Conjunction
  - Instantiation
  - Restatement
    - specification
    - equivalence
    - generalization
  - Alternative
    - conjunctive
    - disjunctive
    - chosen
  - Exception
  - List

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

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?

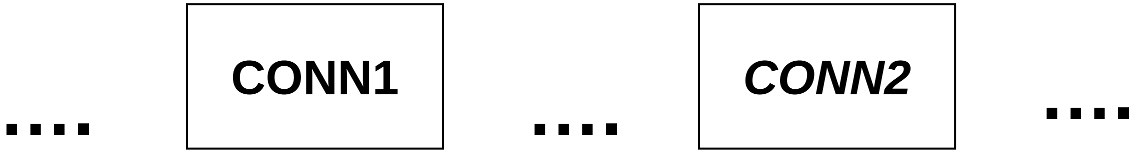
- The annotations does not necessarily lead to a single tree over the entire discourse

  -- comparison with the sentence level

- Complexity of discourse dependencies?

  -- comparison with the sentence level.

Patterns of Consecutive Connectives



…. CONN1 …. *CONN2* ….

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

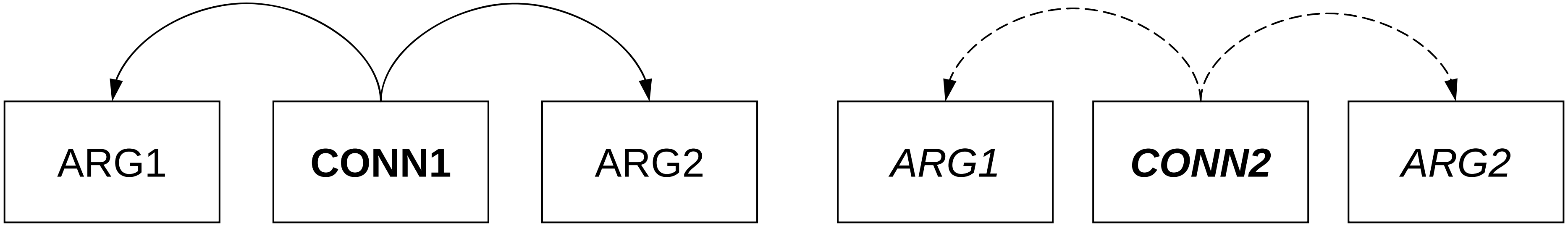
Spans of Consecutive Connectives

- **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



ARG1

CONN1

ARG2

ARG1

CONN2

ARG2

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

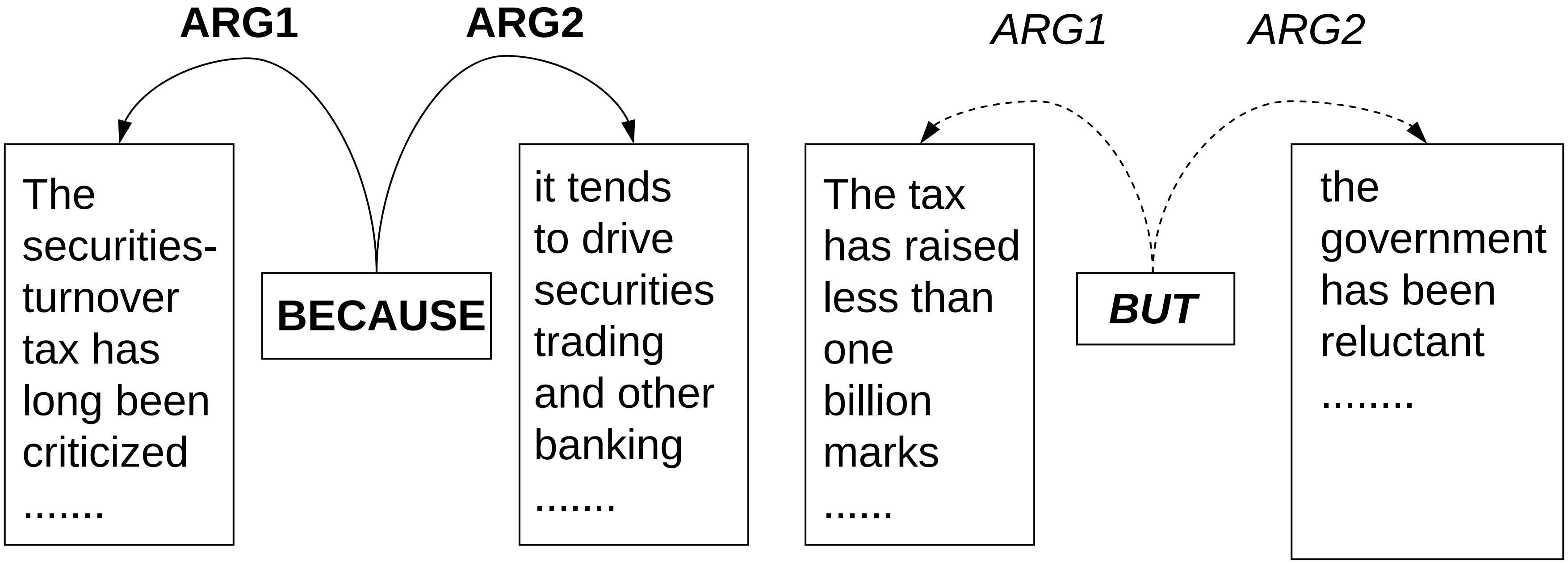
Independent: Example

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 ......

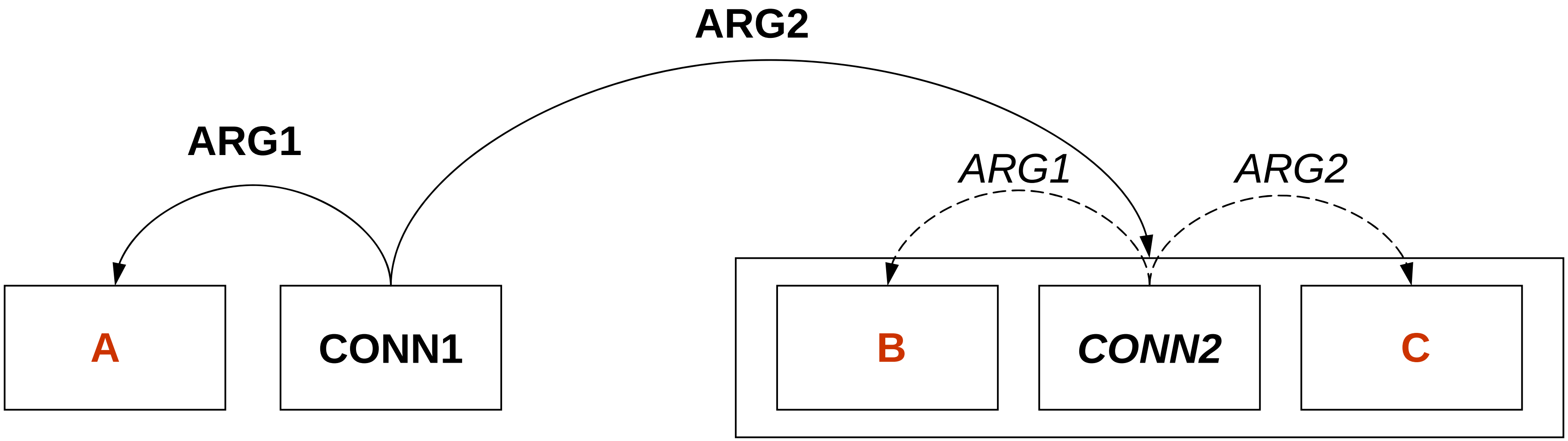
BUT

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

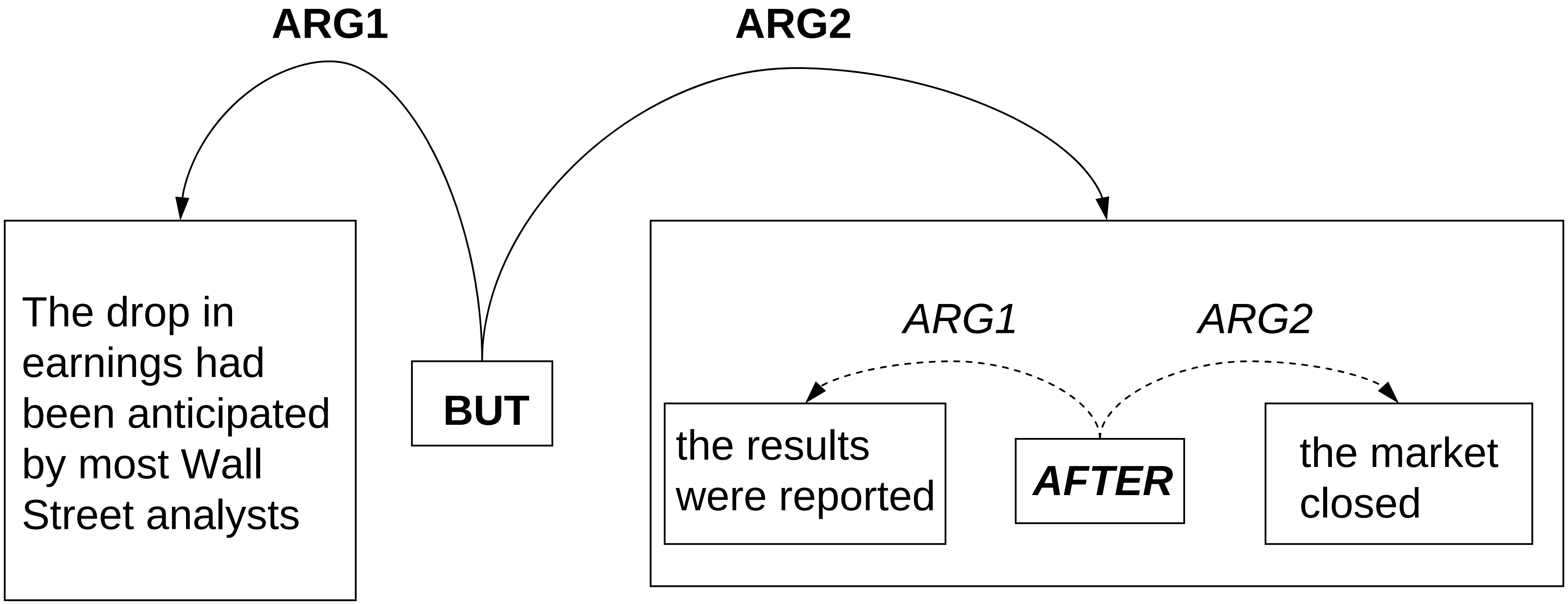
Exhaustively Embedded: Example

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

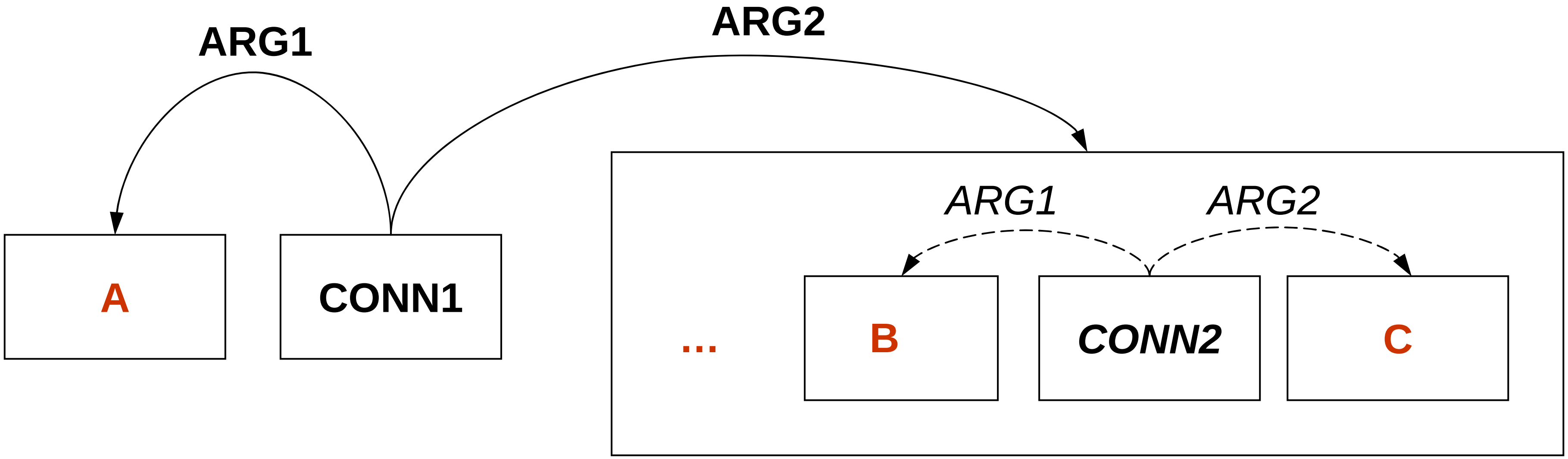
AFTER

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.

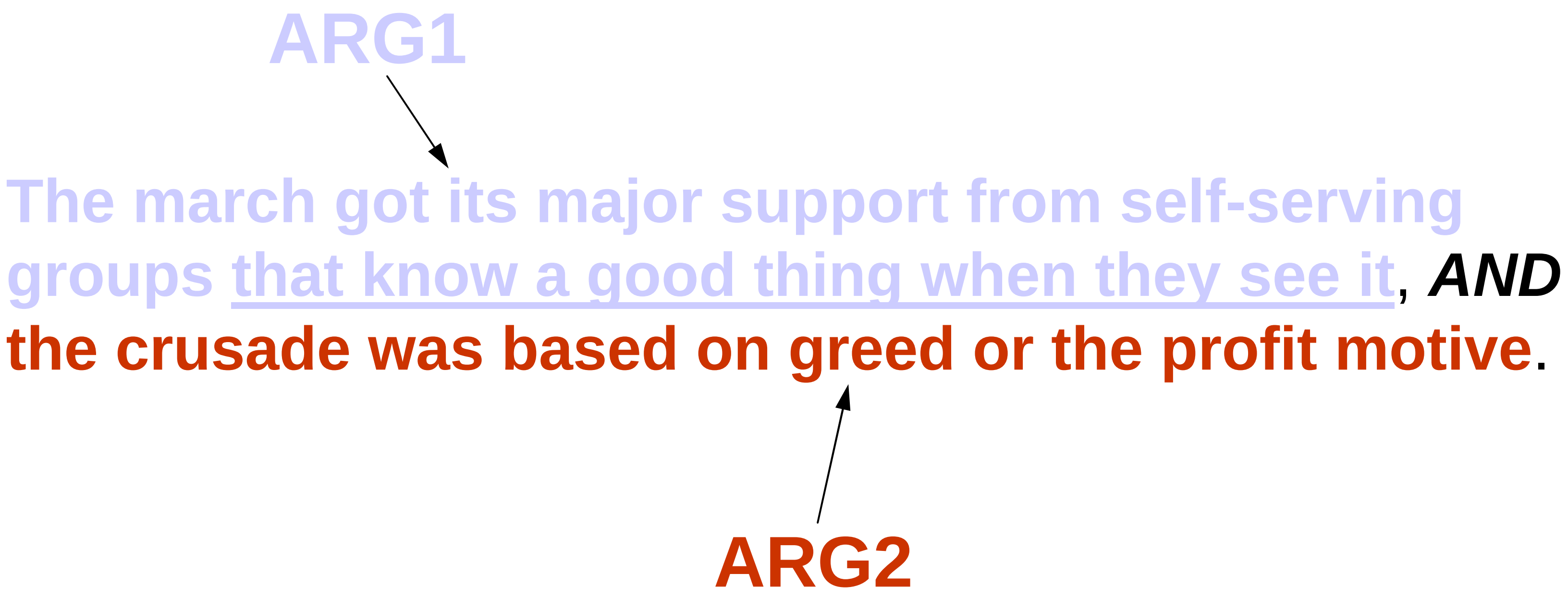
Properly Embedded: Example

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

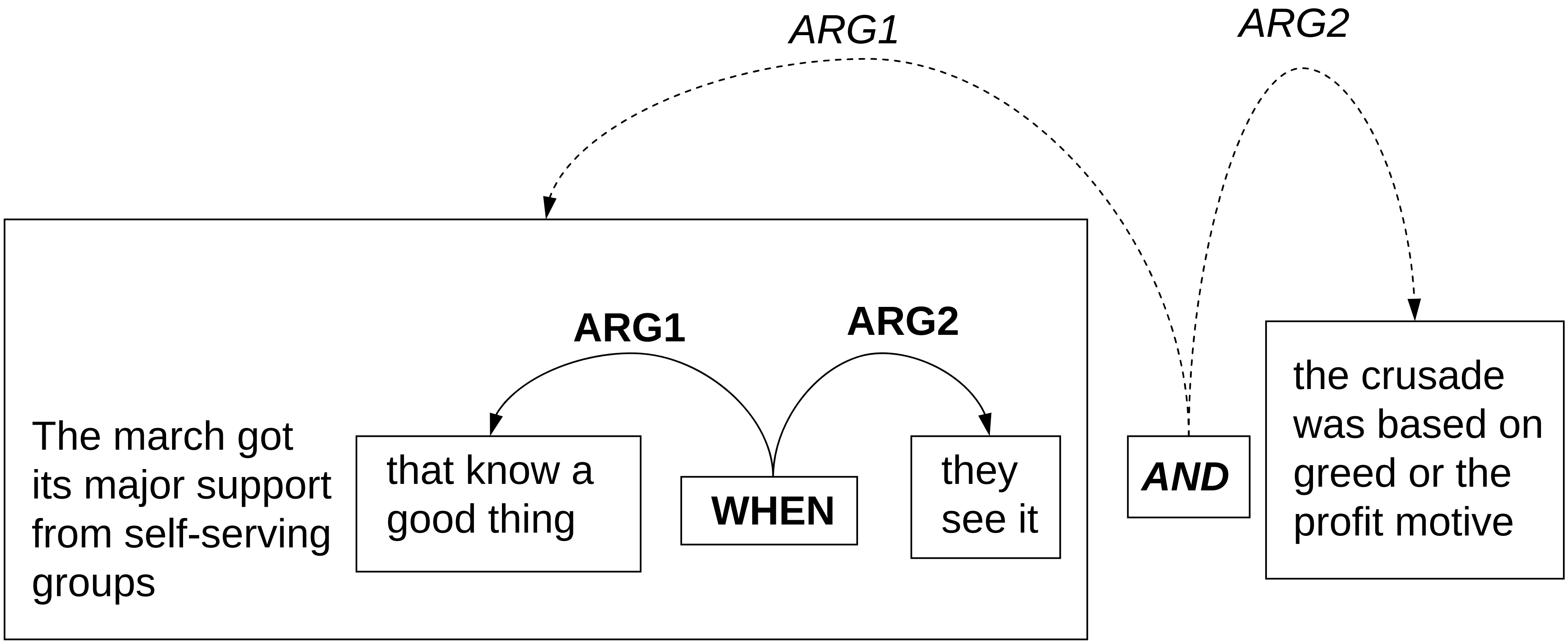


ARG1

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



The march got its major support from self-serving groups

that know a good thing

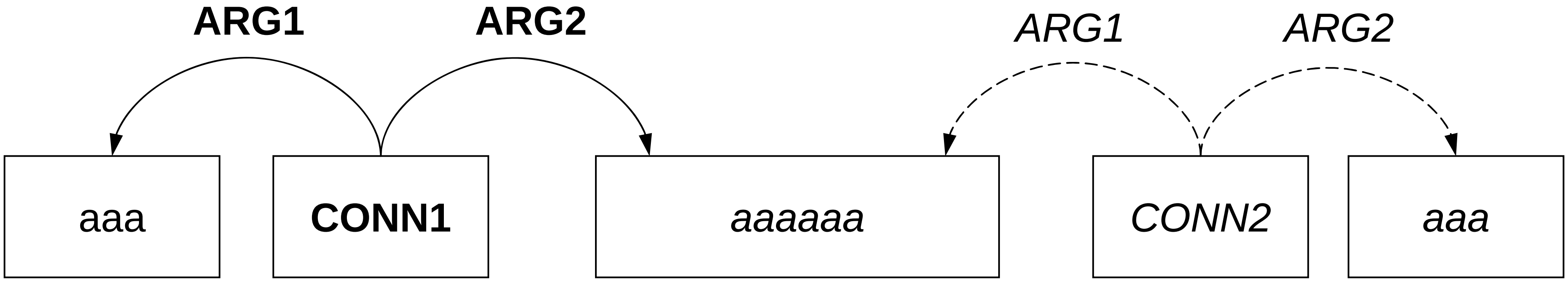
they see it

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

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 and some other debts.

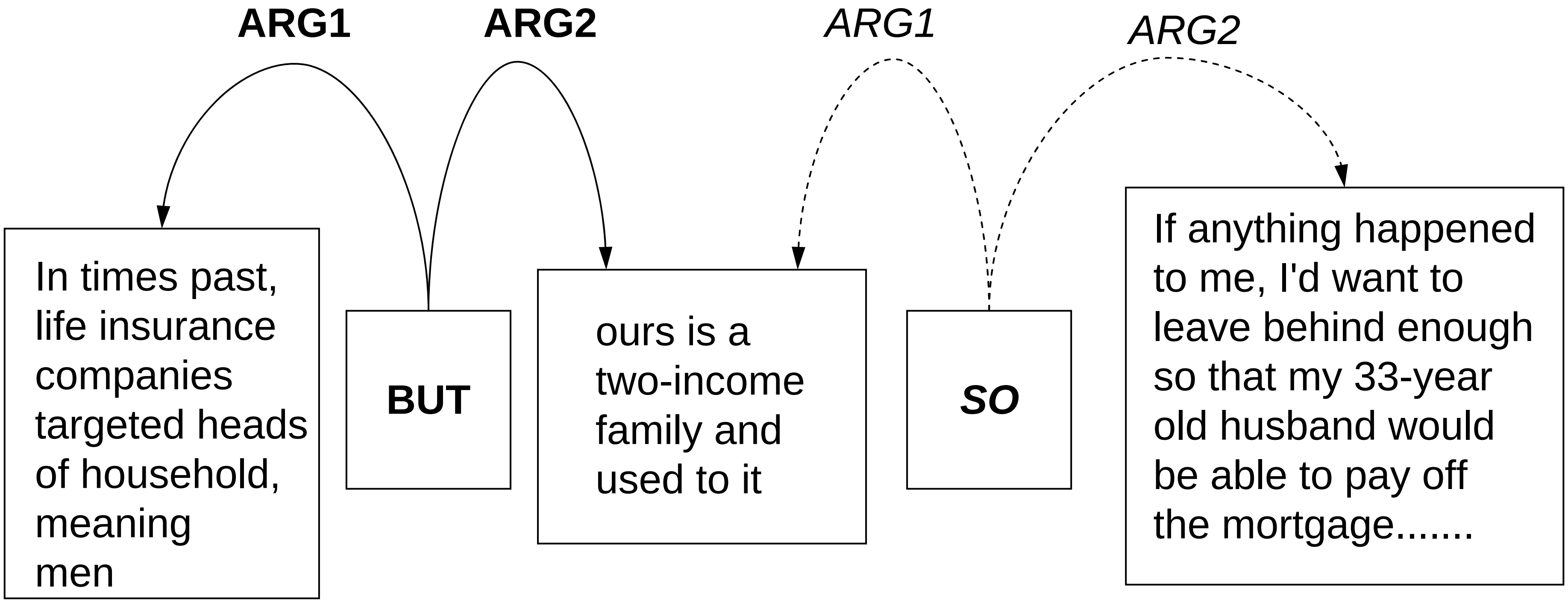
Fully Shared Arg: Example

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

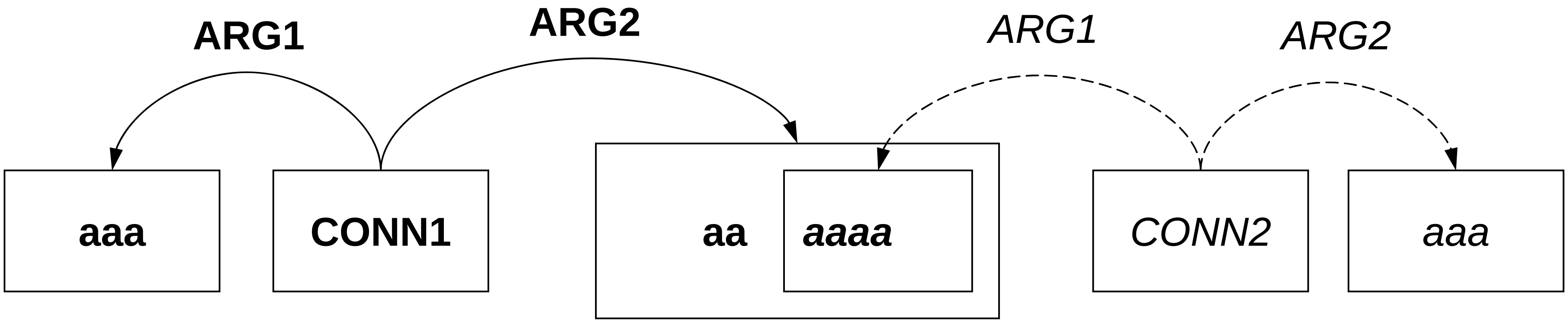
SO

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

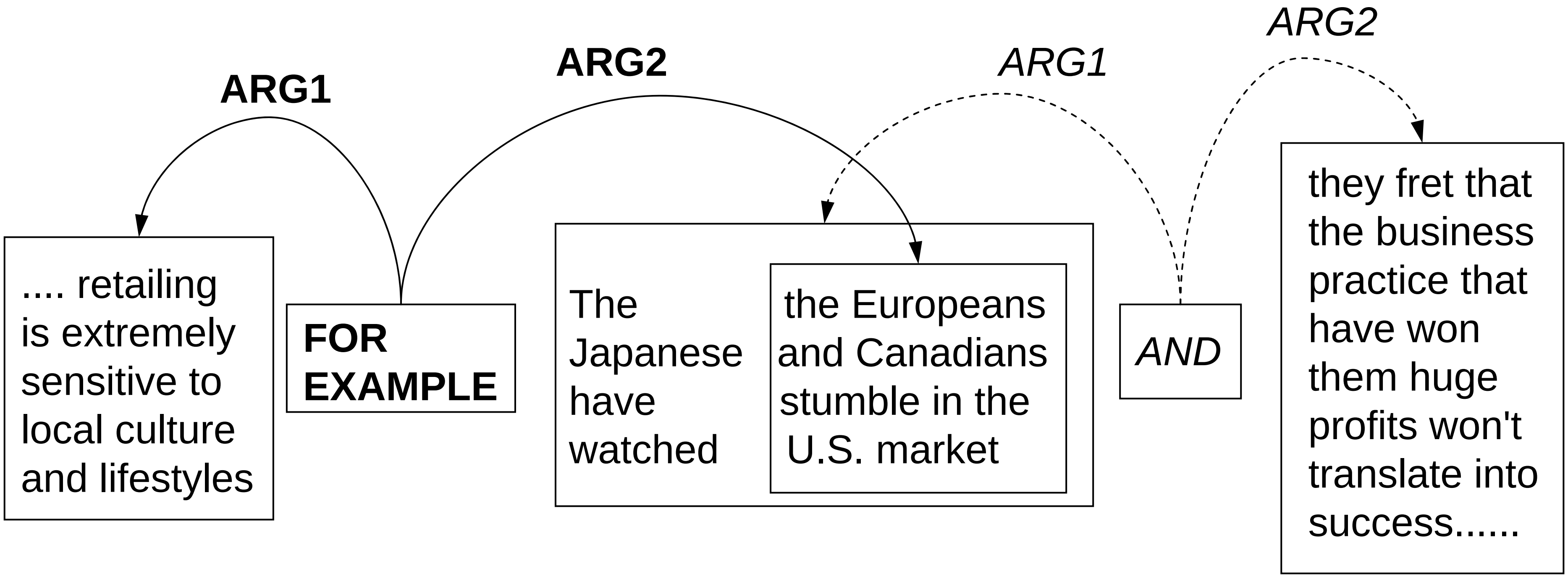
2nd Discourse Relation

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

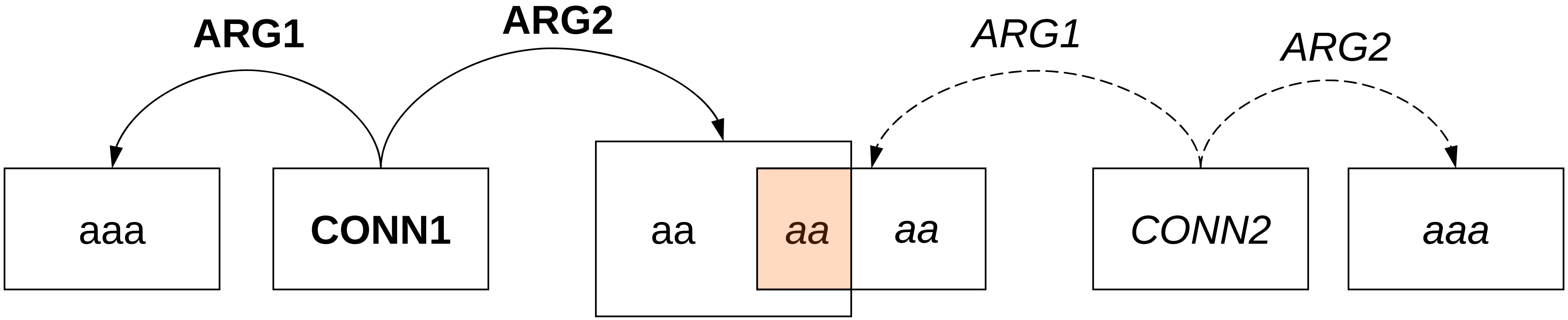
AND

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

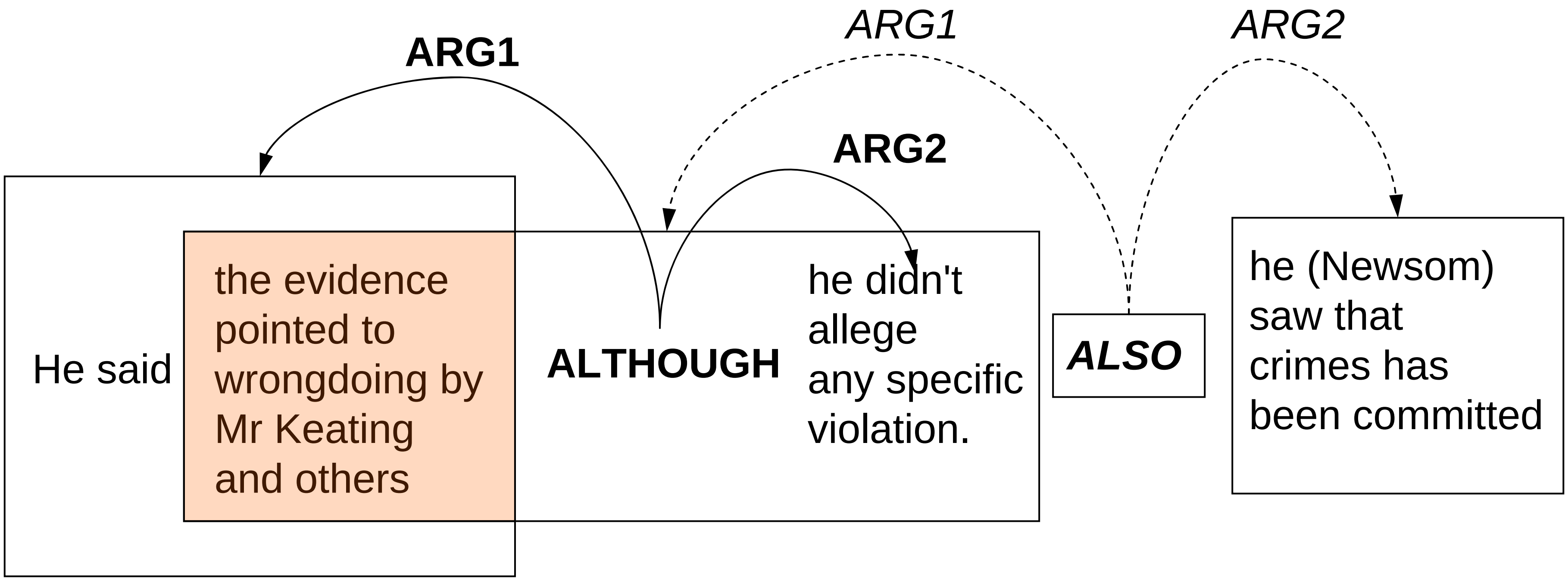
Overlapping Args: Example

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**.

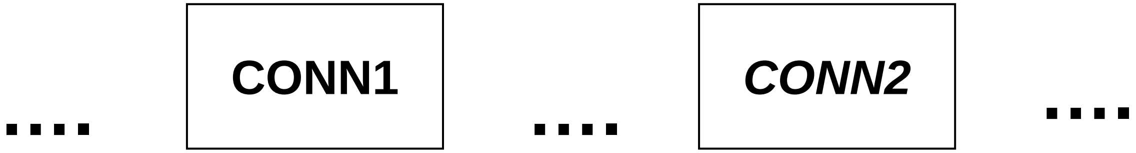
ARG2

Overlapping Args: Example



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

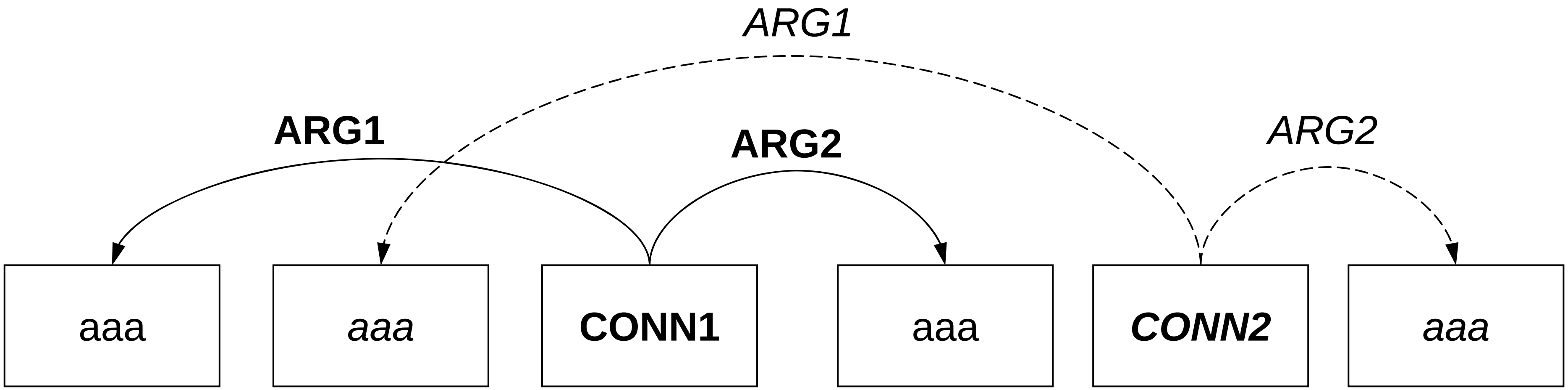


.... CONN1 .... *CONN2* ....

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

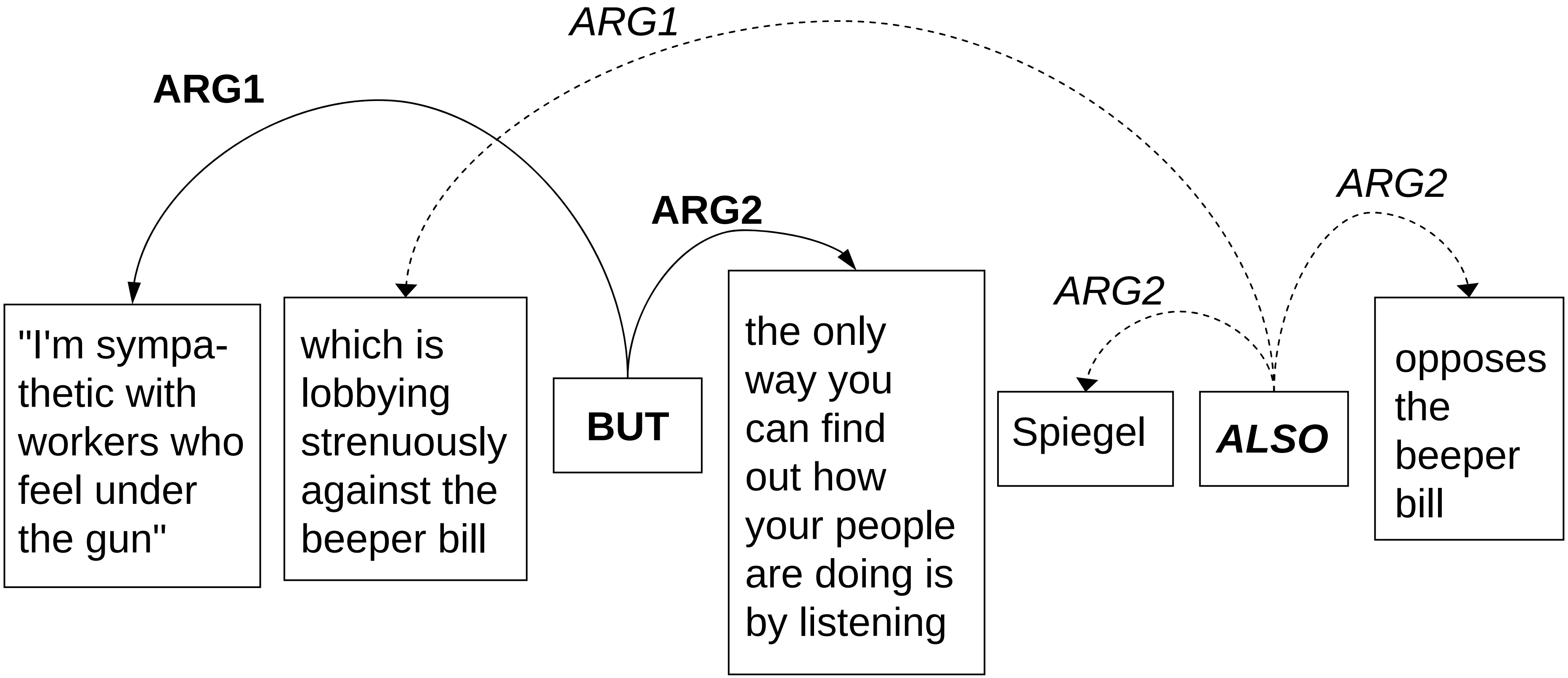
Pure Crossing: Example

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



"I'm sympa-thetic 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

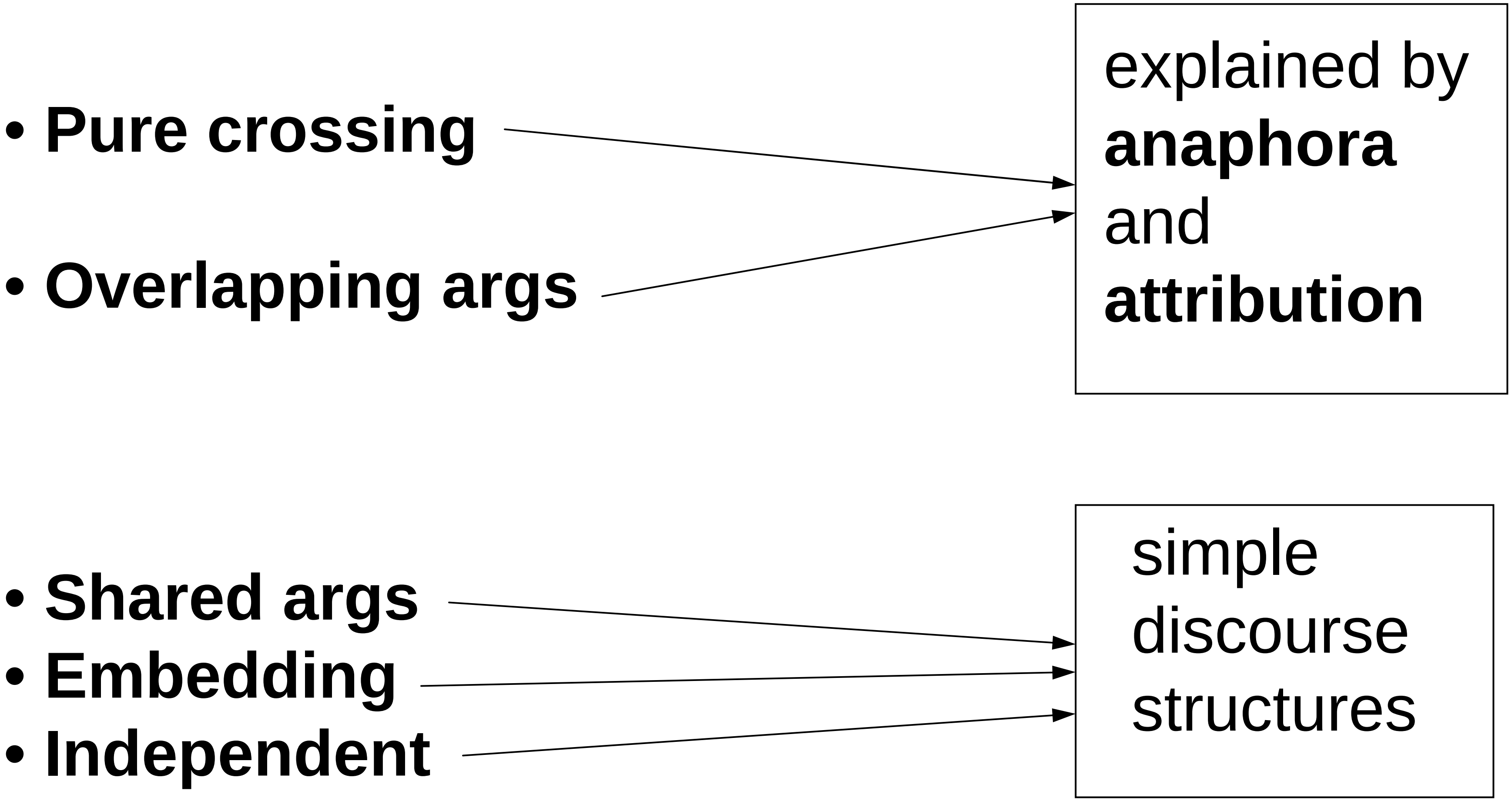
opposes the beeper bill

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

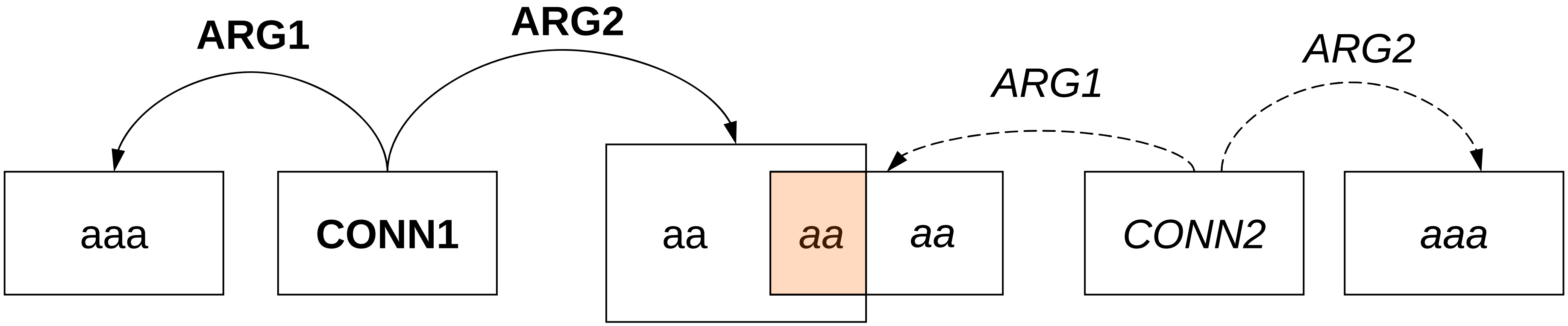
- All cases of pure crossing in the PDTB involve at least one **discourse adverbial**.

- 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.