Discourse Relations and Machine Translation

Bonnie Webber
School of Informatics
University of Edinburgh

December 1, 2012

What are discourse structures?

Discourse structures are the patterns one sees in multi-sentence (multi-clausal) texts.

Recognizing these pattern(s) and what they convey is useful for deriving intended information from the text.

Patterns of Topics

Each element of a topic pattern focuses on one or more entities and their properties. Common to expository text, as in text books, encyclopedias, etc.

<table>
<thead>
<tr>
<th>Wisconsin</th>
<th>Louisiana</th>
<th>Vermont</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Etymology</td>
<td>Etymology</td>
<td>Geography</td>
</tr>
<tr>
<td>2 History</td>
<td>Geography</td>
<td>History</td>
</tr>
<tr>
<td>3 Geography</td>
<td>History</td>
<td>Demographics</td>
</tr>
<tr>
<td>4 Demographics</td>
<td>Demographics</td>
<td>Economy</td>
</tr>
<tr>
<td>5 Law and government</td>
<td>Economy</td>
<td>Law and government</td>
</tr>
<tr>
<td>6 Economy</td>
<td>Law and government</td>
<td>Media</td>
</tr>
<tr>
<td>7 Municipalities</td>
<td>Education</td>
<td>Transportation</td>
</tr>
<tr>
<td>8 ...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Topic Patterns in Wikipedia
Patterns of Functions

Each element of a functional pattern serves a particular purpose with respect to the discourse, as in a news report:

- Headline
- Lead paragraph (sometimes spelled lede), conveying who is involved, what happened, when it happened, where it happened, why it happened, and (optionally) how it happened
- Body, providing more detail about who, what, when, where
- Tail, containing less important information

This is why the first (ie, lead) paragraph is usually the best extractive summary of a news report.

Patterns of Eventualities

Each element describes an event or state, and the individuals involved in them. The pattern structure comes from their spatio-temporal and causal relations.

In late summer 1945, guests are gathered for the wedding reception of Don Vito Corleone’s daughter Connie and Carlo Rizzi. Vito, the head of the Corleone Mafia family, is known to friends and associates as “Godfather.” He and Tom Hagen, the Corleone family lawyer, are hearing requests for favors because, according to tradition, “no Sicilian can refuse a request on his daughter’s wedding day.”

One of the men who asks the Don for a favor is Amerigo Bonasera, a successful mortician and old friend of the Don, whose daughter was beaten by two young men who received minimal punishment. The Don is mostly disappointed in Bonasera, who’d avoided most contact with the Don due to Corleone’s nefarious business dealings.

[http://www.imdb.com/title/tt0068646/synopsis]
Low-level Patterns of Coherence Relations

(1) Come home by 5:00.  $s_1$
   Then we can go to the hardware store before it closes.  $s_2$
   That way we can finish the bookshelves tonight.  $s_3$
   [mp92]

Semantic/Informational  Pragmatic/Intentional

Rhetorical Structure Theory [Mann & Thompson, 1988]

Now consider how textual elements are linked by discourse relations:

Rhetorical Structure Theory (RST):
- Elementary discourse units (EDUs) form a non-overlapping cover of a text.
- Discourse relations hold recursively over discourse elements, with EDUs as terminals.
- Only one relation holds between any two elements, so discourse relations form a tree.
- Some types of relation take one argument (the nucleus) to be more significant to the relation than the other (the satellite), but cf. [ste08].

Rhetorical Structure Theory [Mann & Thompson, 1988]

(2) And the mayor, in an admonition that bears a rhythmic resemblance to Prof. Hill’s, warned that “alcohol leads to betting, which leads to fights.” [wsj_2367]
In the previous example,

(3) Come home by 5:00. \texttt{s1} Then we can go to the hardware store before it closes. \texttt{s2} That way we can finish the bookshelves tonight. \texttt{s3}

only one of the two structures would be assigned to the text.

Discourse relations can hold between discourse segments or groupings (adjacent segments with a common attribution or topic);

Discourse relations can also hold between non-adjacent segments or groupings, or within a grouping.

One discourse relation involves attributing a proposition to an agent.

The resulting structure is a chain graph with crossing links — many from attribution relations.

Discourse relations are signalled by

- an explicit discourse connective: conjunctions and discourse adverbials
- within-paragraph sentence-adjacency, if either an implicit connective can be inferred that expresses the relation between the sentences, or another explicit construction (AltLex) does so.

Each argument of a relation is what is minimally required and sufficient for interpreting it (Minimality Principle).

Must be at least a clause, though possibly a nominalization. (Often a nominalization in Turkish [zey10].)
Following *PropBank*, the argument syntactically attached to the connective is called **Arg2**, and the other argument, **Arg1**.

5. **By most measures, the nation’s industrial sector is now growing very slowly** – if at all. Factory payrolls fell in September. So did the Federal Reserve Board’s industrial-production index. Yet **many economists aren’t predicting that the economy is about to slip into recession**. [wsj_0036]

6. **Mr. Lane’s final purpose isn’t to glamorize the Artist’s vagabond existence. Implicit=RATHER He has a point he wants to make, and he makes it, with a great deal of force.** [wsj_0039]

If annotators felt the relation was **already** expressed, they annotated its lexico-syntactic evidence as an alternative lexicalization or AltLex:

7. **The two companies each produce market pulp, containerboard and white paper. That means goods could be manufactured closer to customers, saving shipping costs**, he said. [wsj_0317]

8. **The new structure would be similar to a recapitalization in which holders get a special dividend yet retain a controlling ownership interest. The difference is that current holders wouldn’t retain majority ownership or control.** [wsj_1531]

EntRel was labelled when an adjacent sentence appeared only related to its predecessor through entity-based coherence.

9. **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.** [wsj_0945]
(10) **Dodge reported an 8% increase in construction contracts awarded in September. NoRel The government counts money as it is spent [wsj_0036]**

Attribution is considered a property of discourse relations and/or their arguments, not a separate discourse relation.

(11) **The administration should now state that if the February election is voided by the Sandinistas, they should call for military aid, said former Assistant Secretary of State Elliot Abrams. “In these circumstances, I think they’d win.” (Wolf & Gibson, Ex. 26, from wsj_0655)**

Discourse relations can share arguments [lee06].

(12) **In times past, life-insurance salesmen targeted heads of household, meaning men, but ours is a two-income family and accustomed 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.**

(13) In times past, life-insurance salesmen targeted heads of household, meaning men, but ours is a two-income family and accustomed 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.
Discourse relations may only provide a partial cover [lee06].

(14) “It’s hard to explain to a 17-year-old why someone they like had to go,” says Mrs. Ward. NoRel Soon, T-shirts appeared in the corridors that carried the school’s familiar red-and-white GHS logo on the front. On the back, the shirts read, “We have all the answers.”

(15) “It’s hard to explain to a 17-year-old why someone they like had to go,” says Mrs. Ward. Soon, T-shirts appeared in the corridors that carried the school’s familiar red-and-white GHS logo on the front. Implicit=WHILE On the back, the shirts read, “We have all the answers.”
Inventories of Discourse Relations

Approaches differ in what semantics (pragmatics) is conveyed by discourse relations and how.

**What sense is conveyed:**
- Definitional approaches, including RST [mt88] and the PDTB [pra08]
- Compositional approaches [san92,kno96,kno98]

**What signals the sense:**
- the discourse connective [san92,kno96,kno98,pra08]
- features throughout the discourse relation [mt88]
- **both** — possibly the same or possibly different senses

Compositional approaches posit a set of features whose values are mutually exclusive, which lead to the sense of a discourse relation and/or discourse connective. [San92]

- **Basic operation:** causal (if a 'relevant' causal connection exists between the arguments) and additive otherwise.
- **Source of coherence:** semantic if the arguments are related in terms of their propositional content, and pragmatic, if in terms of their illocutionary force.
- **Polarity:** positive if the content of the two arguments is linked as they stand, while negative if the content of one is linked to the negation of the content of the other.
- **Order of segments:** basic if the antecedent of a causal relation is on the left, non-basic otherwise.

Definitional approaches define relations in words and/or logic.

**Comparison** applies “when the connective indicates that a discourse relation is established between Arg1 and Arg2 in order to highlight prominent differences between the two situations”. [PDTB 2.0 Annotation Manual]

- **Contrast** applies when a connective indicates that Arg1 and Arg2 share a predicate or property and a difference is highlighted with respect to the values assigned to it.
- **Concession** applies when a connective indicates that one argument describes a situation \( A \) which causes \( C \), while the other asserts (or implies) \( \neg C \). Alternatively, one argument denotes a fact that triggers a set of potential consequences, while the other denies one or more of them.

For Knott [kno96], what signals a discourse relation is an explicit connective.

How it does so is through its compositional features, recognized through **substitutability** tests.

- \( X \) is synonymous with \( Y \) if any context in which one can be used, the other can as well.
- (16) It was a hot day, so therefore they ate dinner outside.
- \( X \) and \( Y \) are exclusive if they can never be substituted for each other in any context.
- (17) I was happy until/* and I became a PhD student.
- (18) I was happy and/* until I wanted other people to be happy as well.
- X is a hypernym of Y if wherever Y can be used, so can X, but not necessarily vice versa. (19) Sally was sick whereas but her brother was healthy. (20) Sally was sick but whereas she went to work anyway.

- X and Y are contingently substitutable if there are some cases where they are substitutable, some cases where X can be used but not Y, and some cases where Y can be used but not X.

For Knott, connective substitutability is a consequence of their features and mutually exclusive values.
- Basic operation: causal vs additive
- Source of coherence: semantic vs pragmatic
- Polarity: positive vs negative
- Order of segments: basic vs non-basic.

Annotated substitutability structure of some English connectives [kno98]

Annotated substitutability structure of some Dutch connectives [kno98]
What signals the sense of a discourse relation

One must take this question seriously:

- the discourse connective
- features in the discourse relation’s arguments.

If the answer is both,

- the same sense may be signalled whether a connective present or not;
- the connective and argument features may possibly signal different senses, both/all of which hold;
- translation may realize in an explicit $L_T$ connective, features conveyed implicitly in $L_S$ arguments;
- translation may even realize in $L_T$ argument features, what was conveyed in an explicit $L_S$ connective;

With a lexicalized view of discourse connectives, the general problems are:

- Given a language, what affixes, words, terms and/or constructions can serve to relate elements in a discourse (i.e., as its discourse connectives)?
- Given a particular token in text, is its current role to relate discourse elements?
- Given a token that does relate discourse elements, which ones does it relate (i.e., which serve as its arguments)?
- Given such a token and its arguments, what sense relation(s) hold between the arguments?

From DConns to their Args: Problem 1

In English, all coordinating and subordinating conjunctions indicate a relation between discourse elements.

- **Coordinating conjunctions** (on clauses or sentences)
  1. (21) Finches eat seeds, and/but/or robins eat worms.
  2. (22) Finches eat seeds. But today, I saw them eating grapes.

- **Subordinating conjunctions**
  1. (23) While finches eat seeds, robins eat worms.
  2. (24) Robins eat worms, just as finches eat seeds.

However, with other parts-of-speech, only certain types indicate a relation between discourse elements — i.e.,

- all **discourse adverbials**
  25. (25) Robins eat both worms and seeds. Consequently they are omnivores.

  26. (26) Robins eat both worms and seeds. Fortunately they prefer worms.

  - Relation between speaker and discourse element
  27. (27) Robins eat both worms and seeds. Frequently they can also be seen eating grubs.

  - Event frequency
Other constructions (alternative lexicalizations) can also serve to indicate a discourse relation \{pra08a, pra:coling10\}:

- **this/that</be> why/when/how</S>**
- **this/that</be> before/after/while/because/if/etc.</S>**
- **the reason/result</be></S>**
- **what’s more</S>, what’s more</Adj></S>**

How can these other constructions be identified automatically?

**Problem 2:** When does an individual token signal a coherence relation, when it may be syntactically ambiguous \{pit09b\}:

(28) Asbestos is harmful **once** it enters the lungs. (subordinating conjunction)

(29) Asbestos was **once** used in cigarette filters. (adverb)

Pitler & Nenkova \{pit09b\} were able to distinguish discourse and non-discourse usage of tokens with an f-score of

- 75.33% based on the string alone
- 88.19% based on syntactic features alone (ignoring the string, but Gold Standard parse)
- 92.28%, using both.
- 94.19%, using interactions between syntactic features as well.

**Problem 3:** Given a token that signals a discourse relation between discourse elements, which ones serve as its arguments?

- **Arg2** – argument syntactically bound to the connective
- **Arg1** – the other argument
With Arg2, the main question is how much is included in the argument— in particular, attribution phrases:

(30) We pretty much have a policy of not commenting on rumors, and I think that falls in that category. [wsj.2314]

(31) 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. [wsj.0098]

With Arg1, the main questions are where it is located (it need not be adjacent to Arg2) and how much is included:

1. Discourse adverbials are anaphoric. Like pronouns, they may refer to an entity introduced earlier in the discourse.

(32) On a level site you can provide a cross pitch to the entire slab by raising one side of the form (step 5, p. 153), but for a 20-foot-wide drive, this results in an awkward inch slant across the drive’s width. Instead, make the drive higher at the center.

2. All parts of a text are not equally essential to an argument:

(33) Big buyers like Procter & Gamble say there are other spots on the globe and in India, where the seed could be grown. "It’s not a crop that can’t be doubled or tripled," says Mr. Krishnamurthy. But no one has made a serious effort to transplant the crop. [wsj.0515]

Here, the quote and its attribution are not essential to the relation headed by But, so can be excluded from Arg1.

Wellner & Pustejovsky [wp07] were the first to try automatically identifying discourse conns and their arguments. They used:

- a “head-based” dependency representation that reduced arg identification to head location;
- a discriminative log-linear ranking model on syntactic, dependency and lexical features, to separately identify connectives and their arguments;
- a log-linear re-ranking model to select the best pair of arguments (Arg1–Arg2), to capture any dependencies between them.
From DConns to their Arguments: Automated processing

<table>
<thead>
<tr>
<th>Type of connective</th>
<th>Ranking Accuracy</th>
<th>Re-ranking Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coordinating conjunctions</td>
<td>75.5%</td>
<td>78.3%</td>
</tr>
<tr>
<td>Subordinating conjunctions</td>
<td>82.2%</td>
<td>86.8%</td>
</tr>
<tr>
<td>Discourse adverbials</td>
<td>42.2%</td>
<td>49%</td>
</tr>
</tbody>
</table>

There are clearly feature-based dependencies between the args of coordinating conjunctions and discourse adverbials.

W&P did not investigate how subordinating conjunctions and their arguments differ from those of CCs and DAs, or whether re-ranking helps some SCs but not others.

Elwell & Baldridge [elw08] show that connective specific models (with the lexical connective as a feature) lead to:
- significantly improvements for discourse adverbials (DAs) from 49.0% to 67.5%
- none for coordinating conjunctions (CCs)
- worse performance for subordinating conjunctions.

Prasad, Webber & Joshi [pra10] show that for inter-sentential connectives — DAs and sentence-initial CCs — location is an important feature:
- If the sentence containing the connective (Arg2) is not paragraph-initial, Arg1 is easier to identify since it is earlier in the paragraph 4301/4373 = 98% of the time.
- On average, there are only 3 sentences in a WSJ paragraph.

Automatic sense labelling of discourse relations

The PDTB takes its sense labels from a 3-level sense hierarchy:

<table>
<thead>
<tr>
<th>Level 1</th>
<th>Level 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal</td>
<td>Synchronous, Precedence, Succession</td>
</tr>
<tr>
<td>Contingency</td>
<td>Cause, Pragmatic cause, Condition, Pragmatic condition</td>
</tr>
<tr>
<td>Comparison</td>
<td>Contrast, Pragmatic contrast, Concession, Pragmatic concession</td>
</tr>
<tr>
<td>Expansion</td>
<td>Conjunction, Instantiation, List, Restatement, Alternative, Exception</td>
</tr>
</tbody>
</table>

Work on Discourse TreeBanks in other languages and genres suggests adding similarity and purpose relations.

Classifying marked sense relations

Some explicit discourse connectives are essentially unambiguous with respect to these senses.

<table>
<thead>
<tr>
<th>Conn</th>
<th>sense</th>
<th>Conn</th>
<th>sense</th>
</tr>
</thead>
<tbody>
<tr>
<td>accordingly</td>
<td>RESULT (5/5)</td>
<td>in addition</td>
<td>CONJUNCTION (165/165)</td>
</tr>
<tr>
<td>additionally</td>
<td>CONJUNCTION (7/7)</td>
<td>moreover</td>
<td>CONJUNCTION (100/101)</td>
</tr>
<tr>
<td>afterward</td>
<td>PRECEDENCE (11/11)</td>
<td>so</td>
<td>RESULT (262/263)</td>
</tr>
<tr>
<td>as a result</td>
<td>RESULT (78/78)</td>
<td>thus</td>
<td>RESULT (112/112)</td>
</tr>
<tr>
<td>consequently</td>
<td>RESULT (10/10)</td>
<td>till</td>
<td>PRECEDENCE (3/3)</td>
</tr>
<tr>
<td>for instance</td>
<td>INSTANTIATION (98/98)</td>
<td>unless</td>
<td>DISJUNCTIVE (94/95)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>instead</td>
<td>CHOSEN ALT (108/112)</td>
</tr>
</tbody>
</table>
Classifying marked sense relations

But several common connectives can express $\geq 1$ sense:

- **since**: CONTINGENCY,CAUSE,REASON (94), TEMPORAL,SUCCESSION (78)
- **as**: TEMPORAL,SYNCHRONY (387), CONTINGENCY,CAUSE,REASON (166)
- **and**: CONTINGENCY,CAUSE,RESULT (38), EXPANSION,CONJUNCTION (2543), both of these simultaneously (138)

[pit09b] get 94.15% accuracy from a simple Naive Bayes classifier on the four **top-level** senses of explicit connectives, using lexical and syntactic features of their arguments, which suffices for these ambiguous connectives.

Classifying unmarked relations

For “unmarked” discourse relations, evidence must be derivable from other features.

(34) [ A high speed train hit a car stopped on a level crossing. ]
   [ The train derailed. ] $\Rightarrow$ **RESULT**

(35) [ The damage to the train was substantial, ]
   [ fortunately nobody was injured ] $\Rightarrow$ **CONTRAST**

(36) **It isn’t just exercise gear that isn’t getting a good workout.** The fitness craze itself has gone soft, the survey found. [wsj_0409] $\Rightarrow$ **CHosen ALTERNATIVE**

(37) **Copper futures prices failed to extend Friday’s rally.** Declines came because of concern that demand for copper may slow down. [wsj_0437] $\Rightarrow$ **CHosen ALTERNATIVE**

Sense relations and discourse chunking

Lin, Ng and Kan [lin10] identify sense relations in the course of end-to-end discourse chunking, which identifies

- Explicit connectives, their arguments and their senses;
- Implicit relations and their senses (for only top 11 sense types, given data sparcity);
- Attribution.

F-score results on gold standard annotation (no error propagation):

- Similar to previous results for each type of explicit connective;
- 40% for implicit connects ($\sim$25-26% with error propagation).

Recent PhD thesis [lin12] describes this work in greater detail.
Discourse Relations in (S)MT

What is being done:
- Work by Thomas Meyer and colleagues at Idiap on disambiguating highly ambiguous explicit discourse connectives prior to translation [mey11,mey12]

What can be done:
- Recognize features needed to map a less specific connective in $L_s$ to its appropriate specific version in $L_T$;
- Handle the discrepancy between the frequency of discourse connectives in texts "born" in a given $L$ and texts that have been translated into $L$;

<table>
<thead>
<tr>
<th></th>
<th>French</th>
<th>German</th>
<th>Dutch</th>
<th>Italian</th>
</tr>
</thead>
<tbody>
<tr>
<td># additions</td>
<td>6</td>
<td>12</td>
<td>19</td>
<td>15</td>
</tr>
<tr>
<td># removals</td>
<td>10</td>
<td>10</td>
<td>7</td>
<td>18</td>
</tr>
</tbody>
</table>

From [zuf12]
- Reformulate sentences based on paraphrasing discourse relations.

References


