Linguistic structure beyond the sentence?

What makes...

▶ An argument persuasive?
▶ A story suspenseful?
▶ A joke funny?

Put another way:

▶ **Grammaticality** is the property that distinguishes well-structured sentences from random sequences of words.
▶ **Coherence** has been proposed to play the same role at the multi-sentence level.

But what are the properties of a coherent text?
This talk

Four models of discourse structure

- Analyses of the same two texts in each model
- Brief summary of current state of corpus annotation and automated discourse parsing
- Applications, especially to machine translation
No, you clearly don’t know who you’re talking to, so let me clue you in. I am not in danger, Skyler. I am the danger. A guy opens his door and gets shot, and you think that of me? No. I am the one who knocks!

https://www.youtube.com/watch?v=3HH9IiHMD2M&t=4s
<table>
<thead>
<tr>
<th>Original</th>
<th>From French</th>
</tr>
</thead>
<tbody>
<tr>
<td>You clearly don’t know who you’re talking to,</td>
<td>You do not know to whom you address.</td>
</tr>
<tr>
<td>so let me clue you in.</td>
<td>Let me be clear.</td>
</tr>
<tr>
<td>I am not in danger, Skyler.</td>
<td>I’m not in danger, Skyler.</td>
</tr>
<tr>
<td>I am the danger.</td>
<td>The danger is me.</td>
</tr>
<tr>
<td>A guy opens his door and gets shot,</td>
<td>A guy gets off at his door.</td>
</tr>
<tr>
<td>and you think that of me?</td>
<td>You see me in his place?</td>
</tr>
<tr>
<td>No. I am the one who knocks!</td>
<td>No. It is I who knocks on the door.</td>
</tr>
</tbody>
</table>
Running example #2

The more people you love, the weaker you are. You’ll do things for them that you know you shouldn’t do. You’ll act the fool to make them happy, to keep them safe. Love no one but your children. On that front, a mother has no choice.

https://www.youtube.com/watch?v=49_cPvbNA54&t=3m47s
The more people you love, the weaker you are.

You’ll do things for them that you know you shouldn’t do.

You’ll act the fool to make them happy, to keep them safe.

Love no one but your children.

On that front, a mother has no choice.

The more you love people, more you are weak.

You will do things knowing that you should not do them.

You will play the madness to make them happy, for their protection.

Love nothing more than your children.

In this world, women have no other choice.
Table of Contents

Rhetorical Structure Theory

The Penn Discourse Treebank

Centering

Lexical Cohesion and Functional Zoning
The projections are in the neighborhood of 50 cents a share to 75 cents, compared with a restated $1.65 a share a year earlier, when profit was $107.8 million on sales of $435.5 million.

Elementary discourse units

RST is built from **elementary discourse units (EDUs)**, which roughly correspond to **clauses**.

- Clauses that are not EDUs: subjects, objects, or complements of a main verb (except attribution verbs).

- “Phrases that begin with a strong discourse marker, such as because, in spite of, as a result of, according to, are treated as EDUs.”

- **Embedded discourse units**: “Relative clauses, nominal postmodifiers, or clauses that break up other legitimate EDUs, are treated as ” (Carlson et al., 2003)

- The RST manual devotes 30 pages to segmentation! But in practice most cases seem to be easy.
You clearly don’t know who you’re talking to,
so let me clue you in.
I am not in danger, Skyler.
I am the danger.
A guy opens his door
and gets shot,
and you think that of me?
No.
I am the one who knocks!
RST segmentation: another example

- The more people you love,
- the weaker you are.
- You’ll do things for them that you know you shouldn’t do.
- You’ll act the fool
- to make them happy,
- to keep them safe.
- Love no one but your children.
- On that front, a mother has no choice.
Discourse relations and nuclearity

Discourse units are built up through discourse relations.

\[ \text{ANTITHESIS} \]

[I am not in danger, Skyler.]\(^1\text{A}\) [I am the danger.]\(^1\text{B}\)

1B is the nucleus, 1A is the satellite.

- “In general, nuclear units can be understood by themselves, in isolation of the satellite units that they refer to.” (see strong compositionality criterion)
- Similar idea to dependency grammar; “right frontier constraint” \( \approx \) projectivity.
Multinuclear relations

Discourse relations can involve multiple nuclei.

[You’ll act the fool]\(^2\text{A}\) [to keep them happy,]\(^2\text{B}\) [to keep them safe.]\(^2\text{C}\)

- “A multinuclear relation contains two or more units or spans of equal importance in the discourse.”
- “Nuclearity assignment is often determined simultaneously with the rhetorical relation. What counts as a nucleus... can rarely be determined in isolation.”

Jacob Eisenstein: Models of Discourse Structure
Types of discourse relations

There are 78 types of discourse relations!

- “Presentational relations” are about persuading the reader to accept the nucleus.
- “Subject matter relations” reflect an underlying semantic relationship between the related units.
- List of RST relations: http://www.sfu.ca/rst/01intro/definitions.html.
Definitions that follow are quoted from there.
Non-volitional cause

“The more people you love,]^{3A}
[the weaker you are.]^{3B}

- “S, by means other than motivating a volitional action, caused N; without the presentation of S, R might not know the particular cause of the situation... N is more central than S.”

- In NON-VOLITIONAL RESULT, N causes S.
Evidence

[The more people you love,]^{4A} [the weaker you are.]^{4B} [You’ll do things for them that you know you shouldn’t do.]^{4C}

▶ “R’s comprehending S increases R’s belief of N”
The more people you love, the weaker you are. You’ll act the fool to make them happy, to keep them safe.\(^5\)A [Love no one but your children.\(^5\)B [On that front, a mother has no choice.\(^5\)C

- Motivation: “Comprehending S increases R’s desire to perform action in N”
- Concession: “W acknowledges a potential or apparent incompatibility between N and S.”
Inter-annotator agreement

<table>
<thead>
<tr>
<th>Span</th>
<th>Nuclearity</th>
<th>Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>88.7</td>
<td>77.7</td>
<td>65.8</td>
</tr>
</tbody>
</table>

- **Span** refers to both the EDU boundaries and the boundaries of composite discourse units.
- **Nuclearity** requires getting the span right, and then identifying the correct nucleus.
- **Relation** requires getting the nucleus right, and then identifying the correct discourse relation.
EDU segmentation

Generative model of Soricut & Marcu (2003):

Best discriminative result: 91.0% F1, using CRF + Reranker (Xuan Bach et al., 2012).
Generative RST parsing

\[ S_1 = 1 \]
\[ P_S ( [2,2,3] \mid (2, SBAR) < (1, VP), (3, S) < (2, VP) ) = 0.47 \]
\[ P_T ( \text{ENABLEMENT} \rightarrow \text{NS} \mid S(\text{to}) < VP(\text{use}) ) = 0.88 \]
\[ S_2 = P_S \cdot P_T = 0.40 \]

ATtribution

\[ P_S ( [1,1,3] \mid (2, SBAR) < (1, VP), (3, S) < (2, VP) ) = 0.37 \]
\[ P_T ( \text{ATTRIBUTION} \rightarrow \text{SN} \mid SBAR(\text{will}) < VP(\text{says}) ) = 0.009 \]
\[ \text{Score} = S_1 \cdot S_2 \cdot P_S \cdot P_T = 0.001 \]
Modern RST parsing

Approaches:

▶ Classification + CKY (Hernault et al., 2010; Feng & Hirst, 2012)
▶ Sequence labeling (Ghosh et al., 2012; Joty et al., 2013)
▶ Shift-reduce (Sagae, 2009; Ji & Eisenstein, 2014; Heilman & Sagae, 2015)
▶ Representation learning (Ji & Eisenstein, 2014; Li et al., 2014)

Best results (on gold EDU segments)

<table>
<thead>
<tr>
<th></th>
<th>Span</th>
<th>Nuclearity</th>
<th>Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feng &amp; Hirst (2014)</td>
<td>85.7</td>
<td>71.0</td>
<td>58.2</td>
</tr>
<tr>
<td>Li et al. (2014)</td>
<td>82.9</td>
<td>73.0</td>
<td>60.6</td>
</tr>
<tr>
<td>Ji &amp; Eisenstein (2014)</td>
<td>81.6</td>
<td>71.0</td>
<td>61.8</td>
</tr>
</tbody>
</table>
Applications of RST parsing

- Summarization (Marcu, 1999)
- Question-answering (Jansen et al., 2014)
- Sentiment analysis (Chardon et al., 2013; Bhatia et al., 2015; Hogenboom et al., 2015)
- Machine translation
  - Discourse tree alignment (Marcu et al., 2000)
  - Chinese-English tree-to-string (Tu et al., 2013), based on “discourse chunking” Sporleder & Lapata (2005).
- Evaluation (Guzmán et al., 2014)
Annotated corpora

- **RST Treebank**: 385 English newswire documents
- **Multilingual RST Treebank**: 15 parallel technological abstracts, in English, Spanish, and Basque
- **CSTNews Corpus**: 50 documents in Brazilian Portuguese
- **SFU Review Corpus**: English and Spanish, 400 review documents each
- **Potsdam Commentary Corpus**: German newstext (no longer available?)
Discourse graphs

Wolf & Gibson (2005) argue that many discourses cannot be fully described by a tree, and that graphs are more appropriate.

0. Farm prices in October edged up 0.7% from September
1. as raw milk prices continued to rise,
2. the Agriculture Department said.
3. Milk sold to the nations dairy plants and dealers averaged $14.50 for each hundred pounds,
4. up 50 percent from September and up $1.50 from October 1988,
5. the department said.

(see http://www.isi.edu/~marcu/discourse/Discourse%20structures.htm)
Table of Contents

Rhetorical Structure Theory

The Penn Discourse Treebank

Centering

Lexical Cohesion and Functional Zoning
AltLex

Ms. Bartlett’s previous work, which earned her an international reputation in the non-horticultural art world, often took gardens as its nominal subject.

Mayhap this metaphorical connection made the BPC Fine Arts Committee think she had a literal green thumb

Source: 09/wsj_0084

D-LTAG: Lexically anchored discourse relations

- Choosing from the 78 RST relations is hard!
- Let’s just assume relations \( \approx \) connectives (e.g., if, so, however, . . .)
  - If the relation is **explicitly marked** by a connective, then the annotator must specify the spans of the arguments.
  - For all adjacent sentences in which there is no connective, the annotator should specify the **implicit relation**, if any.
From parsing to chunking

RST assumes discourse forms a rooted tree... is that realistic?

[Prasad et al., 2014]

PDTB parsing is sometimes called “shallow discourse parsing” (Prasad et al., 2010).
Relation taxonomy

**TEMPORAL**
- Asynchronous
- Synchronous
  - precedence
  - succession

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

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

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

Jacob Eisenstein: Models of Discourse Structure
Explicit discourse relations

No, [you clearly don’t know who you’re talking to]_{a1}, so [let me clue you in]_{a2}.

The annotator or parser must specify:

1. the sense of the discourse relation
2. the spans of the two arguments
   - Arguments can be less than or more than a clause,
   - can cross sentence boundaries,
   - need not be adjacent to the connective (“no restriction on how far an argument can be from its connective”, pg 11),
   - and can even contain each other!
Implicit discourse relations

For every pair of sentences, specify either:

- a connective that could fit between them;
- or that there is a relation that is already lexicalized by an expression that is not a connective, \textit{AltLex};
- or that the sentences are related by shared entities, \textit{EntRel};
- or that there is no relation, \textit{NoRel};
Example 1

- You clearly don’t know who you’re talking to, so let me clue you in.
- (?) I am not in danger, Skyler.
- (?) I am the danger.
- (?) A guy opens his door and gets shot, and you think that of me?
- No, I am the one who knocks!
Example 1

- You clearly don’t know who you’re talking to, so let me clue you in.
- (Specifically,) I am not in danger, Skyler.
- (Instead,) I am the danger.
- (For example,) A guy opens his door and gets shot, and you think that of me?
- No, I am the one who knocks!
Example 1

- You clearly don’t know who you’re talking to, so let me clue you in.

- (Expansion.Instantiation) I am not in danger, Skyler.

- (Contrast.Opposition) I am the danger.

- (Expansion.Instantiation) A guy opens his door and gets shot, and you think that of me?

- (Contrast.Opposition) I am the one who knocks!
Example 2

- The more people you love, the weaker you are.
- (?) You’ll do things for them that you know you shouldn’t do.
- (?) You’ll act the fool to make them happy, to keep them safe.
- (?) Love no one but your children.
- (?) On that front, a mother has no choice.
Example 2

- The more people you love, the weaker you are.
- (For example,) You’ll do things for them that you know you shouldn’t do.
- (In addition,) You’ll act the fool to make them happy, to keep them safe.
- (Therefore,) Love no one but your children.
- On that front (ALTLEX), a mother has no choice.
Example 2

▶ The more people you love, the weaker you are.
▶ (Expansion.Instantiation) You’ll do things for them that you know you shouldn’t do.
▶ (Expansion.Restatement) You’ll act the fool to make them happy, to keep them safe.
▶ (Contingency.PragmaticCause) Love no one but your children.
▶ [Contingency.Cause.Reason] a mother has no choice.
Interannotator agreement

Argument spans:
- 90.2% exact match
- 94.5% “partial match.”

Sense annotation:
- Class: 94%
- Type: 84%
- Subtype: 80%

(Prasad et al., 2008)
Datasets

From Prasad et al. (2014):

<table>
<thead>
<tr>
<th>Name</th>
<th>Coverage</th>
<th># relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDTB</td>
<td>WSJ news, essays</td>
<td>40,600</td>
</tr>
<tr>
<td>BioDRB</td>
<td>Biomed papers</td>
<td>5,859</td>
</tr>
<tr>
<td>LADTB</td>
<td>Arabic news</td>
<td>6,328</td>
</tr>
<tr>
<td>Chinese DTB</td>
<td>Xinhua news</td>
<td>3,951</td>
</tr>
<tr>
<td>Turkish DB</td>
<td>novels, news, etc.</td>
<td>8,484</td>
</tr>
<tr>
<td>Hindi DRB</td>
<td>news</td>
<td>5000ish</td>
</tr>
<tr>
<td>PDT 3.0</td>
<td>news</td>
<td>20,542</td>
</tr>
</tbody>
</table>
PDTB parsing

End-to-end
- Classifier pipeline (Lin et al., 2010, 2014), achieving an “overall F1” of 38.2%.
- No published work since then! But CoNLL 2015 shared task results due soon...

Explicit discourse relations
- Classifying the sense of explicit discourse relations is not difficult (Pitler et al., 2008).
Implicit relation classification

- Lexical features, Levin verb classes, polarity (Pitler et al., 2009)
- Selection and aggregation of bilexical features (Park & Cardie, 2012; Biran & McKeown, 2013)
- Coreference and Brown clustering (Rutherford & Xue, 2014)
- Recursive neural networks (Ji & Eisenstein, 2015), obtaining 44.6% accuracy on level-2 discourse relations.

(Evaluation in this area is kind of a mess.)
Applications of PDTB-style annotation

- **Text quality assessment** (Lin et al., 2011), including MT output (Li et al., 2014)
- **Sentiment and opinion analysis** (Somasundaran et al., 2008)
- **Translation of explicit connectives after sense classification** (Meyer et al., 2012, 2015)

In general, applications have focused on explicit discourse relations, maybe because implicit relation classification is currently so bad.
Table of Contents

Rhetorical Structure Theory

The Penn Discourse Treebank

Centering

Lexical Cohesion and Functional Zoning

- Hypothesis: coherent discourses use referring expressions and syntactic position to support inference over references.

- Goal: model reader’s *attentional state* over discourse entities.
Centering theory: definitions

Utterance:
A sequence of words (typically a sentence or clause) at a particular point in a discourse.

The centers of an utterance:
Entities (semantic objects) which link the utterance to the previous and following utterances.
Centering: assumptions

In each utterance, **some discourse entities are more salient** than others.

We maintain a **list of discourse entities, ranked by salience**. The position in this list determines **how easy it is to refer back to an entity** in the next utterance. Each utterance updates this list.

This list is called the **local attentional state**.
The two centers of an utterance

Backward-looking: mentioned in $U_i$ and $U_{i-1}$

Forward-looking: mentioned in $U_i$

The **forward looking center** of an utterance $U_n$ is a partially ordered **list of the entities** mentioned in $U_n$. The ordering reflects **salience** within $U_n$:

\[
\text{subject} > \text{direct object} > \text{object}, \ldots.
\]

The **backward looking center** of an utterance $U_n$ is the highest ranked **entity** in the forward looking center of $U_{n-1}$ that is mentioned in $U_n$. 
Center realization and pronouns

Observation: Only the most salient entities of $U_{n-1}$ can be referred to by pronouns in $U_n$.

**Constraint/Rule 1:**

If *any* element of $FW(U_{n-1})$ is realized as a *pronoun* in $U_n$, then the $BW(U_n)$ has to be realized as a *pronoun* in $U_n$ as well.

*Sue told Joe to feed her dog.*  
$BW=$Susan,  $FW=$\{Sue, Joe, dog\}

*He asked her what to feed it.*  
$BW=$Sue,  $FW=$\{Joe, Sue, dog\}

✔ Constraint obeyed

*He asked Sue what to feed it.*  
$BW=$Sue,  $FW=$\{Joe, Sue, dog\}

✗ Constraint violated:  
*Sue* should be a pronoun as well.
Transitions between sentences

Center continuation:

\[ BW(U_n) = BW(U_{n-1}). \text{ BW}(U_n) \text{ is highest ranked element in } FW(U_n) \]

- Sue gave Joe a dog.
- She told him to feed it well.
- She asked him whether he liked the gift.

\[ BW=Sue, \; FW=\{Sue, Joe, dog\} \]
\[ BW=Sue, \; FW=\{Sue, Joe, gift\} \]

Center retaining:

\[ BW(S_n) = BW(S_{n-1}). \text{ BW}(S_n) \neq \text{ highest ranked element in } FW(S_n) \]

- Sue gave Joe a dog.
- She told him to feed it well.
- John asked her what to feed him.

\[ BW=Sue, \; FW=\{Sue, Joe, dog\} \]
\[ BW=Sue, \; FW=\{Joe, Sue, dog\} \]

Center shifting:

\[ BW(S_n) \neq BW(S_{n-1}) \]

- Susan gave Joe a dog.
- She told him to feed it well.
- The dog was very cute.

\[ BW=Sue, \; FW=\{Sue, Joe, dog\} \]
\[ BW=dog, \; FW=\{dog\} \]
Local coherence: preferred transitions

Rule/Constraint 2:

- Center *continuation* is preferred over center *retaining*.
- Center *retaining* is preferred over center shifting.

Local coherence is achieved by maximizing the number of center continuations.
## Example

<table>
<thead>
<tr>
<th>$U_n$</th>
<th>$C_f$</th>
<th>$C_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>You don’t know who you’re talking to, so let me clue you in.</td>
<td>Skyler, Walter</td>
<td></td>
</tr>
<tr>
<td>I am not in danger, Skyler</td>
<td>Walter, danger, Skyler</td>
<td>Skyler</td>
</tr>
<tr>
<td>I am the danger.</td>
<td>Walter, danger</td>
<td>Walter</td>
</tr>
<tr>
<td>A guy opens his door and gets shot, and you think that of me?</td>
<td>a guy, the door</td>
<td></td>
</tr>
<tr>
<td>No. I am the one who knocks!</td>
<td>Skyler, Walter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Walter</td>
<td>Walter</td>
</tr>
</tbody>
</table>

Jacob Eisenstein: Models of Discourse Structure
Instantiating centering theory

Poesio et al. (2004) argue that:

- Centering is underspecified (politely: “a parametric theory”).
- Its predictions depend on precise definitions of utterance and reference.

Barzilay & Lapata (2008) propose the Entity Grid, which is more amenable to computation on real texts.

- The entity grid preserves the key ideas of centering: it focuses on the syntactic role of entities in adjacent utterances.
- Its parameters can be learned from data.
You don’t know who you’re talking to, so let me clue you in. I am not in danger, Skyler. I am the danger. A guy opens his door and gets shot, and you think that of me? No. I am the one who knocks!

<table>
<thead>
<tr>
<th></th>
<th>Skyler</th>
<th>Walter</th>
<th>danger</th>
<th>a guy</th>
<th>the door</th>
</tr>
</thead>
<tbody>
<tr>
<td>You don’t know who you’re talking to, so let me clue you in.</td>
<td>S</td>
<td>- (O?)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>I am not in danger, Skyler.</td>
<td>X</td>
<td>S</td>
<td>X</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>I am the danger.</td>
<td>-</td>
<td>S</td>
<td>O</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A guy opens his door and gets shot, and you think that of me?</td>
<td>S</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>No. I am the one who knocks!</td>
<td>-</td>
<td>S</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Skyler</td>
<td>Walter</td>
<td>danger</td>
<td>a guy</td>
<td>the door</td>
</tr>
<tr>
<td>----------------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>-------</td>
<td>----------</td>
</tr>
<tr>
<td>You do not know to</td>
<td>S</td>
<td>- (O?)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>whom you address.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Let me be clear.</td>
<td>- (S?)</td>
<td>O</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>I am not in danger,</td>
<td>X</td>
<td>S</td>
<td>X</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Skyler.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The danger is me.</td>
<td>-</td>
<td>O</td>
<td>S</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A guy gets off at his</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>S</td>
<td>O</td>
</tr>
<tr>
<td>door,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>you see me in his</td>
<td>S</td>
<td>O</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>place?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No! It is I who knocks</td>
<td>-</td>
<td>S</td>
<td>-</td>
<td>-</td>
<td>X</td>
</tr>
<tr>
<td>on the door</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Entity grid example 2

<table>
<thead>
<tr>
<th></th>
<th>you</th>
<th>people you love</th>
</tr>
</thead>
<tbody>
<tr>
<td>The more people you love, the weaker you are</td>
<td>S</td>
<td>X</td>
</tr>
<tr>
<td>You’ll do things for them that you know you shouldn’t do</td>
<td>S</td>
<td>X</td>
</tr>
<tr>
<td>You’ll act the fool to make them happy, to keep them safe</td>
<td>S</td>
<td>O</td>
</tr>
<tr>
<td>Love no one but your children</td>
<td>X</td>
<td>-</td>
</tr>
<tr>
<td>On that front, a mother has no choice</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Jacob Eisenstein: Models of Discourse Structure
<table>
<thead>
<tr>
<th>Sentence</th>
<th>Label</th>
<th>Entity Grid</th>
</tr>
</thead>
<tbody>
<tr>
<td>The more you love people, more you are weak.</td>
<td></td>
<td>S O</td>
</tr>
<tr>
<td>You will do things knowing that you should not do them.</td>
<td></td>
<td>S -</td>
</tr>
<tr>
<td>You will play the madness to make them happy, for their protection.</td>
<td></td>
<td>S O</td>
</tr>
<tr>
<td>Love nothing more than your children.</td>
<td></td>
<td>X -</td>
</tr>
<tr>
<td>In this world, women have no other choice.</td>
<td></td>
<td>- -</td>
</tr>
</tbody>
</table>

(To be fair, most of the “gaps” in these entity grids are missing in the French too.)
Applications

Centering theory
- Original intent was largely for text generation (Kibble, 1999; Karamanis et al., 2009)
- Pronoun resolution (Manabu & Kouji, 1996; Kong et al., 2009)

The entity grid
- Readability prediction Pitler & Nenkova (2008)
- Essay scoring Burstein et al. (2010)
- Thread disentanglement Elsner & Charniak (2010)
Referring expressions: is syntax enough?

Some examples from Lascarides & Asher (2007)

1. A man walked in. He ordered a beer.
2. Every man walked in. ?? He ordered a beer.
3. John can open Bill’s safe. He’s going to have to get the combination changed soon.¹

L&A see these examples as demonstrating the necessity of framing discourse structure in terms of dynamic semantics, motivating Segmented Discourse Representation Theory (SDRT).

¹L&A attribute this example to Hobbs (1985), which I can’t find.
Table of Contents

Rhetorical Structure Theory

The Penn Discourse Treebank

Centering

Lexical Cohesion and Functional Zoning
Cohesion in English

- **Coherence** is a property of well-structured texts.
- Halliday & Hasan (1976) define **cohesion** as the set of linguistic devices that create coherence.
Cohesive devices

Conjunction  You clearly don’t know who you’re talking to, so let me clue you in.

Reference  The more people you love, the weaker you are. You’ll do things for them that you know you shouldn’t.

Ellipsis  A guy opens his door and (he) gets shot...

Substitution  On that front, a mother has no choice.

Lexical cohesion  I am not in danger. I am the danger. A guy opens his door... I am the one who knocks.
Lexical cohesion and discourse structure

Halliday and Hassan didn’t believe in “discourse structure,” per se!²

Whatever relation there is among the parts of a text the sentences, the paragraphs, or turns in a dialogue it is not the same as structure in the usual sense, the relation which links the parts of a sentence or a clause. [pg. 6]

Between sentences, there are no structural relations. [pg. 27]

²Thanks to Joshi, Prasad, and Webber (2006) for reminding me of these quotes.
Hearst (1997) showed that lexical cohesion could be quantified and used to support unsupervised topic segmentation.
Argumentative zoning

▶ BKG: General scientific background (yellow)
▶ OTH: Neutral descriptions of other people's work (orange)
▶ OWN: Neutral descriptions of own, new work (blue)
▶ AIM: Statements of the particular aim of the current paper (pink)
▶ TXT: Statements of textual organization of the current paper (red)
▶ CTR: Contrastive or comparative statements about other work; explicit mention of weaknesses of other work (green)
▶ BAS: Statements that own work is based on other work (purple)
Some genres with conventionalized functional organization:

- **research papers**: abstract, background, methods, results, discussion
- **inverted pyramid**: lede paragraph, body, tail
- **thesis, antithesis, synthesis**

Recognizing functional patterns could improve information extraction in conventionalized domains:

- **biomedical abstracts** (McKnight & Srinivasan, 2003)
- **legal documents** (Palau & Moens, 2009)
Conventionalized topic structures

Wikipedia articles about US states

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

Chen et al. (2009) used probability distributions over permutations to model conventionalized topic sequences.
Cohesion in machine translation

Lexical cohesion provides a measure of document-level translation quality that is relatively easy to compute.

- Wong & Kit (2012) incorporate cohesion into an MT evaluation metric...

- but Carpuat & Simard (2012) show that MT output is often too lexically consistent.

- “Lexical cohesion trigger” for phrase-based MT (Ben et al., 2013)

- Encourage the generation of more cohesive devices in the target (Tu et al., 2014)

- Document-level scoring to improve overall lexical cohesion (Ding et al., 2014)
Models of discourse relations:

- **RST**: informational + intentional relations; tree-structured, full coverage.

- **PDTB**: relations are lexically anchored; organization is shallow and cross-cutting.

- **Centering**: relations are based on information status of entities; adjacent units only.

- **Cohesion**: cohesive devices between sentences; can be abstracted into discourse segments or functional zones.
Discourse relations and semantics

- Parsing discourse relations is hard, largely because semantic understanding is required.
- But the connection to semantics is exactly why these relations are promising for MT.
- Discourse relations are a lot easier to annotate than open-domain semantics.
- Maybe we should focus on predicting these relations, and let “representation learning” do the hard work of modeling meaning.
References


References


Prasad, R., Joshi, A. K., & Webber, B. L. (2010). Exploiting scope for shallow discourse parsing. In LREC.


