Discourse structure and coherence

Christopher Potts

CS 244U: Natural language understanding
Mar 1
Discourse segmentation and discourse coherence

1. **Discourse segmentation**: chunking texts into coherent units. (Also: chunking separate documents)

2. **(Local) discourse coherence**: characterizing the meaning relationships between clauses in text.
Discourse segmentation examples

- Most Newsworthy Info

- Important Details

- Other General Info
  - Background Info

(The inverted pyramid design)
Clinical Comparison of Full-Field Digital Mammography and Screen-Film Mammography for Detection of Breast Cancer

John M. Lewin¹, Carl J. D’Orsi², R. Edward Hendrick¹, Lawrence J. Moss², Pamela K. Isaacs¹, Andrew Karellas² and Gary R. Cutter⁴

¹ University of Colorado Health Sciences Center, 4200 E. 9th Ave., Mail Stop F724, Denver, CO 80262.
² University of Massachusetts Medical Center, 55 Lake Ave. N., Worcester, MA 01655.
³ Northwestern University Medical School, 357 E. Chicago Ave., Chicago, IL 60611.
⁴ AMC Cancer Research Center, 1800 Pierce St., Lakewood, CO 80232.

OBJECTIVE. The purpose of this work is to compare full-field digital mammography and screen-film mammography for the detection of breast cancer in a screening population.

SUBJECTS AND METHODS. Full-field digital mammography was performed in addition to screen-film mammography in 6736 examinations of women 40 years old and older presenting for screening mammography at either of two institutions. Two views of each breast were acquired with each technique. The digital and screen-film mammograms were each interpreted independently. In addition to a clinical assessment, each finding was assigned a probability of malignancy for use in receiver operating characteristic analysis. In cases in which the digital and screen-film interpretations differed, a side-by-side analysis was performed to determine the reasons for the discrepancy. With few exceptions, findings detected on either technique were evaluated with additional imaging and, if warranted, biopsy.

RESULTS. Additional evaluation was recommended on at least one technique in 1467 cases. These additional evaluations led to 181 biopsies and the detection of 42 cancers. Nine cancers were detected only on digital mammography, 15 were detected only on screen-film mammography, and 18 were detected on both. The difference in cancer detection is not statistically significant (p > 0.1). Digital mammography resulted in fewer recalls than did screenfilm mammography (799 vs 1007, p < 0.001). The difference between the receiver operating characteristic curve area for digital (0.74) and screen-film (0.80) mammography was not significant (p > 0.1). Reasons for discrepant interpretations of cancer were approximately equally distributed among those relating to lesion conspicuity, lesion appearance, and interpretation.

CONCLUSION. No significant difference in cancer detection was observed between digital mammography and screen-film mammography. Digital mammography resulted in fewer recalls than did screen-film mammography.
Discourse segmentation examples

Identification of Genes Required for the Function of Non-Race-Specific mlo Resistance to Powdery Mildew in Barley

A. Freidenhoven, C. Peterhansel, J. Kurth, F. Kreuzaler and P. Schulze-Lefert
Rheinisch-Westfälische Technische Hochschule Aachen, Department of Biology 1, Woringer Weg 1, D-52074 Aachen, Germany

Recessive alleles (mlo) of the Mlo locus in barley mediate a broad, non-race-specific resistance reaction to the powdery mildew fungus Erysiphe graminis f sp hordei. A mutational approach was used to identify genes that are required for the function of mlo. Six susceptible M2 individuals were isolated after inoculation with the fungal isolate K1 from chemically mutagenized seed carrying the mlo-5 allele. Susceptibility in each of these individuals is due to monogenic, recessively inherited mutations in loci unlinked to mlo. The mutants identify two unlinked complementation groups, designated Ror1 and Ror2 (required for mlo-specified resistance). Both Ror genes are required for the function of different tested mlo alleles and for mlo function after challenge with different isolates of E. g. f sp hordei. A quantitative cytological time course analysis revealed that the host cell penetration efficiency in the mutants is intermediate compared with mlo-resistant and Mlo-susceptible genotypes. Ror1 and Ror2 mutants could be differentiated from each other by the same criterion. The spontaneous formation of cell wall appositions in mlo plants, a subcellular structure believed to represent part of the mlo defense, is suppressed in mlo/lor genotypes. In contrast, accumulation of major structural components in the appositions is seemingly unaltered. We conclude that there is a regulatory function for the Ror genes in mlo-specified resistance and propose a model in which the Mlo wild-type allele functions as a negative regulator and the Ror genes act as positive regulators of a non-race-specific resistance response.

(Pubmed less structured abstract)
Discourse segmentation examples

38 of 44 people found the following review helpful:

**Move over, Robert Jordan.**, July 19, 1998

By A Customer

This review is from: A Game of Thrones (A Song of Ice and Fire, Book 1) (Mass Market Paperback)

As a fantasy reader of somewhat high standards, I have always had a proclivity for "epic" fantasy. Nothing else really satisfies my desire for an absorbing story. George R.R. Martin has, with this book, taken the field dominated by such giants as Jordan, Williams, and Kay and blown a great big gust of fresh air into it. Not only does this book have the complicated plot and intricate character development that is common to these three talented authors, but it has a certain brutal realism to it. Granted, we're talking about an invented realm, but never before in all the books that I have read has any author taken his portrayal of all the brutality of human nature to this level. Part of what makes Jordan, Williams, and Kay so brilliant is that they write *human* characters, and good and bad are rarely well delineated. What sets Martin apart is his sheer, brutal, mind-numbing honesty. He doesn't pull any punches, and neither do any of his characters. This is life, in all its pain and glory. Honor is not as important as we would like it to be, and things do not all go well as long as we wish for it hard enough. Here, there is no destructive force stronger than the power of men. There is no evil greater than that in the hearts of men. And there is no power, once man has decided to destroy, that can stop him. **This novel is a masterpiece; beautifully crafted, shockingly realistic, and a joy to read.** However, don't expect to come out of reading this with your ideals intact.

Help other customers find the most helpful reviews

Was this review helpful to you?

(5-star Amazon review)
Discourse segmentation examples

41 out of 50 people found the following review helpful:

What's left unsaid, February 12, 2004

By A Customer

Amazon Verified Purchase (What's this?)

This review is from: A Game of Thrones (A Song of Ice and Fire, Book 1) (Mass Market Paperback)

All of the other excellent reviews of this series are correct. The writing is wonderful. The characters are real. The plot is intricate, fascinating, and never predictable. Et cetera. But none of the reviewers complained about the one thing that has led me to stop reading after plugging through the first two books: This is the darkest, bleakest, most depressing book I have ever read! You must never, ever let yourself bond with a hero, a good, kind, strong, resourceful person who in a 'normal' book would win a gratifying victory at the end of the book. This is because chances are your hero will soon die, most likely brutally. Most (eventually all???) of the good guys die in this book! And everyone is always having to look over his shoulder to see which one of his supposed friends is plotting his death. Innocent children are brutally murdered and their heads put up on pikes. Innocent peasants are slowly hanged, kicking, their eyes bulging out. Their rescuers, instead of pulling off a valiant rescue, are themselves captured and tortured. There are innumerable rapes, including several fairly explicit portrayals of vicious gang rapes of peasant women by invading troops. Every time I finished a reading session I felt depressed. I've never seen so much plague, betrayal, death, and destruction in a novel. It's unrelenting. I don't care how wonderful the writing is. I simply couldn't take it anymore. I want to be uplifted by a book, made to smile and feel vicariously triumphant. I don't want to be beaten down and defeated over and over and over. I had to stop reading.

Help other customers find the most helpful reviews

Was this review helpful to you?  Report abuse  Permalink

Comments (2)

(3-star Amazon review)
Discourse segmentation applications (complete in class)

- mtg summary - find dec.
- general making
- doc sim at the level of structure
- parsing technical doc
- WSD improvements
- background + foreground - task
  - task
  - news
Coherence examples

1. Sam brushed his teeth. He got into bed. He felt a certain ennui.
2. Sue was feeling ill. She decided to stay home from work.
4. The senator introduced a new initiative. He hoped to please undecided voters.
5. Linguists like quantifiers. In his lectures, Richard talked only about every and most.
6. In his lectures, Richard talked only about every and most. Linguists like quantifiers.
Coherence examples

1. Sam brushed his teeth. then He got into bed. then He felt a certain ennui.
2. Sue was feeling ill. so She decided to stay home from work.
3. Sue likes bananas. but Jill does not.
4. The senator introduced a new initiative. because He hoped to please undecided voters.
5. Linguists like quantifiers. for example In his lectures, Richard talked only about every and most.
6. In his lectures, Richard talked only about every and most. in general Linguists like quantifiers.
Coherence examples

1. Sam brushed his teeth. then He got into bed. then He felt a certain ennui.
2. Sue was feeling ill. so She decided to stay home from work.
3. Sue likes bananas. but Jill does not.
4. The senator introduced a new initiative. because He hoped to please undecided voters.
5. Linguists like quantifiers. for example In his lectures, Richard talked only about every and most.
6. In his lectures, Richard talked only about every and most. in general Linguists like quantifiers.
7. A: Sue isn’t here.  
   B: She is feeling ill.
8. A: Where is Bill?  
   B: In Bytes Café.
   B: Here you go. (Stone 2002)
Coherence examples

1. Sam brushed his teeth. then He got into bed. then He felt a certain ennui.
2. Sue was feeling ill. so She decided to stay home from work.
3. Sue likes bananas. but Jill does not.
4. The senator introduced a new initiative. because He hoped to please undecided voters.
5. Linguists like quantifiers. for example In his lectures, Richard talked only about every and most.
6. In his lectures, Richard talked only about every and most. in general Linguists like quantifiers.
7. A: Sue isn’t here.  
   B: because She is feeling ill.
8. A: Where is Bill?  
   B: answer In Bytes Café.
   B: fulfillment Here you go. (Stone 2002)
Coherence in linguistics

Extremely important sub-area:

- Driving force behind coreference resolution (Kehler et al. 2007).
- Driving force behind the licensing conditions on ellipsis (Kehler 2000, 2002).
- Alternative strand of explanation for the inferences that are often treated as conversational implicatures in Gricean pragmatics (Hobbs 1979).
- Motivation for viewing meaning as a dynamic, discourse-level phenomenon (Asher and Lascarides 2003).

For an overview of topics, results, and theories, see Kehler 2004.
<table>
<thead>
<tr>
<th>Overview</th>
<th>Discourse segmentation</th>
<th>Discourse coherence theories</th>
<th>Penn Discourse Treebank 2.0</th>
<th>Unsupervised coherence</th>
<th>Conclusion</th>
</tr>
</thead>
</table>

Coherence applications in NLP (complete in class)
Plan and goals

Plan

- Unsupervised and supervised discourse segmentation
- Discourse coherence theories
- Introduction to the Penn Discourse Treebank 2.0
- Unsupervised discovery of coherence relations

Goals

- **Discourse segmentation**: practical, easy to implement algorithms that can improve lots of information extraction tasks.
- **Discourse coherence**: a deep, important, challenging task that has to be solved if we are to achieve robust NLU
Overview

Discourse segmentation

Discourse coherence theories

Penn Discourse Treebank 2.0

Unsupervised coherence

Conclusion

Discourse segmentation

Clinical Comparison of Full-Field Digital Mammography and Screen-Film Mammography for Detection of Breast Cancer

John M. Lewin1, Carl J. D’Orsi2, R. Edward Hendrick1,1, Lawrence J. Moses3, Pamela K. Isaac4, Andrew Kerlias5 and Gary R. Cudney6

1University of Colorado Health Sciences Center, 4200 E. 9th Ave., Mail Stop P274, Denver, CO 80262
2University of Massachusetts Medical Center, 65 Lake Ave. N., Worcester, MA 01655
3Northwestern University Medical School, 601 E. Chicago Ave., Chicago, IL 60611
4AMC Cancer Research Center, 1600 Pierers St., Lakewood, CO 80222.

OBJECTIVE. The purpose of this work is to compare full-field digital mammography and screen-film mammography for the detection of breast cancer in a screening population.

SUBJECTS AND METHODS. Full-field digital mammography was performed in addition to screen-film mammography in 6716 examinations of women 40 years old and older presenting for screening mammography at either of two institutions. Two views of each breast were acquired with each technique. The digital and screen-film mammograms were each interpreted independently. In addition to a clinical assessment, each finding was assigned a probability of malignancy for use in receiver operating characteristic analysis. In cases in which the digital and screen-film interpretations differed, a side-by-side analysis was performed to determine the reasons for the discrepancy. With few exceptions, findings detected on either technique were evaluated with additional imaging and, if warranted, biopsy.

RESULTS. Additional evaluation was recommended in at least one technique in 1467 cases. Three additional evaluations led to 181 biopsies and the detection of 42 cancers. Nine cancers were detected only on digital mammography, 15 were detected only on screen-film mammography, and 18 were detected on both. The difference in cancer detection is not statistically significant (p = 0.1). Digital mammography resulted in fewer recalls than did screen-film mammography (79% vs. 100%, p = 0.001). The difference between the receiver operating characteristic curve area for digital (0.74) and screen-film (0.80) mammography was not significant (p = 0.1). Reasons for discordant interpretations of cancer were approximately equally distributed among those relating to lesion complexity, lesion appearance, and interpretation.

CONCLUSION. No significant difference in cancer detection was observed between digital mammography and screen-film mammography. Digital mammography resulted in fewer recalls than did screen-film mammography.

Identification of Genes Required for the Function of Non-Race-Specific m(6)A Resistance to Powdered Mildew in Barley

A. Freiheitshoven, C. Peterhansl, J. Karth, F. Kresse and P. Schiena-Lofert

Blackhead Institute, Technical University of Munich, Department of Biology’s University, Munich, Germany

Resistant alleles (m(6)A) of the m(6)A locus in barley mediate a broad, non-race-specific resistance reaction to the powdery mildew fungi Erysiphe graminis f.sp. hordei. A mutational approach was used to identify genes that are required for the function of m(6)A. Six susceptible Mj individuals were isolated after inoculation with the fungal isolate K1 from genetically outcrossed wheat carrying the m(6)A allele. Susceptibility to all of these individuals is due to monogenic, recessive inherited mutations in loci underlying the m(6)A. Mutants the variety (K5 derivatives) were used to identify lesions needed for resistance and lesion function after challenge with different isolates of E. g. f.sp. hordei. A quantitative semi-physiological assay revealed that the host cell peroxidation efficiency in the mutants is intermediate between m(6)A-resistant and Mj-susceptible heterozygotes. Both Mj-1 and Mj-2 genes are required for the function of different race-sensitive m(6)A alleles and for lesion function after challenge with different isolates of E. g. f.sp. hordei. A semi-quantitative transgenic expression assay revealed that the host cell peroxidation efficiency in the mutants is intermediate between m(6)A-resistant and Mj-susceptible heterozygotes. In contrast, accumulation of major structural components in the pathogen is not significantly reduced. Our conclusion that there is a reaction to the m(6)A genes act as positive regulators of a non-race-specific resistance response.

38 of 44 people found the following review helpful:

Move over, Robert Jordan., July 19, 1998

By A Customer

This review is from: A Game of Thrones (A Song of Ice and Fire, Book 1) (Mass Market Paperback)

As a fantasy reader of somewhat high standards, I have always had a productivity for epic fantasy. Nothing else really satisfies my desire for an absorbing story. George R.R. Martin has, with this book, taken the field dominated by such giants as Jordan, Williams, and Kay and blown a great big gust of fresh air into it. Not only does this book have the complicated plot and intricate character development that is common to these three talented authors, but it has a certain brutal realism to it. Granted, we're talking about an invented realm, but never before in all the books that I have read has any author taken his portrayal of all the brutality of human nature to this level. Part of what makes Jordan, Williams, and Kay so brilliant is that they write *human* characters, and good and evil are rarely well delineated. What sets Martin apart is his sheer, brutal, mind-numbing honesty. He doesn't pull any punches, and neither do any of his characters. This is life, in all its glory and pain. Honor is not as important as we would like it to be, and things do not all go well as long as we wish for it hard enough. Here, there is no destructive force stronger than the power of men. There is no evil greater than that in the hearts of men. And there is no power, once man has decided to destroy, that can stop him. This novel is a masterpiece: beautifully crafted, shockingly realistic, and a joy to read. However, don't expect to come out reading this with your ideals intact.

Help other customers find the most helpful reviews

Was this review helpful to you?

41 of 50 people found the following review helpful:

What's left unsaid., February 12, 2004

By A Customer

Amazon Verified Purchase (What's this?)

This review is from: A Game of Thrones (A Song of Ice and Fire, Book 1) (Mass Market Paperback)

All of the other excellent reviews of these series are correct. The writing is wonderful. The characters are real. The plot is intricate, fascinating, and never predictable. Etcetera. But none of the reviewers complained about the one thing that has led me to stop reading after plugging through the first two books: This is the darkest, bleakest, most depressing book I have ever read! You must never, ever let yourself bond with a hero, a good, kind, strong, resourceful person who in a normal book would win a gratifying victory at the end of the book. This is because chances are your hero will soon die, most likely brutally. Most (eventually all???) of the good guys die in this book! And everyone is always having to look over his shoulder to see which one of his supposed friends is plotting his death. Innocent children are brutally murdered and their heads put up on pikes. Innocent peasants are slowly hanged, kicked, their eyes bulging out. Their rescuers, instead of pulling off a valiant rescue, are themselves captured and tortured. There are innumerable rapes, including several fairly explicit portrayals of vicious gang rapes of peasant women by invading troops. Every time I finished a reading session I felt sad. I've never seen so much death, betrayal, death, and destruction in a novel. It's unrelenting. I don't care how wonderful the writing is. I simply couldn't take it anymore. I want to be uplifted by a book: made to smile and feel vicariously triumphant. I don't want to be beaten down and defeated over and over and over. I had to stop reading.

Help other customers find the most helpful reviews

Was this review helpful to you?

Report abuse | Permalink

Comments (2)
Discourse segmentation

Hearst’s 21-paragraph science news article *Stargazer*

![Figure 5](image_url)

Judgments of seven readers on the *Stargazer* text. Internal numbers indicate location of gaps between paragraphs; $x$-axis indicates token-sequence gap number, $y$-axis indicates judge number, a break in a horizontal line indicates a judge-specified segment break.

1—3 Intro – the search for life in space
4—5 The moon’s chemical composition
6—8 How early earth-moon proximity shaped the moon
9—12 How the moon helped life evolve on earth
13 Improbability of the earth-moon system
14—16 Binary/trinary star systems make life unlikely
17—18 The low probability of nonbinary/trinary systems
19—20 Properties of earth’s sun that facilitate life
21 Summary
The TextTiling algorithm (Hearst 1994, 1997)

<table>
<thead>
<tr>
<th>s₁</th>
<th>sum [ s₁ ]</th>
<th>sum [ s₁ ]</th>
<th>sum [ s₁ ]</th>
<th>…</th>
</tr>
</thead>
<tbody>
<tr>
<td>s₂</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s₃</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s₄</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s₅</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s₆</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s₇</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Score this boundary via cosine similarity between the blocks’ vectors

Score vector S: \( b_{1,2} \)
The TextTiling algorithm (Hearst 1994, 1997)

Score vector $S$: $b_{1,2}$ $b_{2,3}$
The TextTiling algorithm (Hearst 1994, 1997)

<table>
<thead>
<tr>
<th>( w_1 )</th>
<th>( w_2 )</th>
<th>( w_3 )</th>
<th>( \cdots )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s_1 )</td>
<td>( s_1 )</td>
<td>( s_1 )</td>
<td>( \cdots )</td>
</tr>
<tr>
<td>( s_2 )</td>
<td>( s_2 )</td>
<td>( s_2 )</td>
<td>( \cdots )</td>
</tr>
<tr>
<td>( s_3 )</td>
<td>( s_3 )</td>
<td>( s_3 )</td>
<td>( \cdots )</td>
</tr>
</tbody>
</table>

Score this boundary via cosine similarity between the blocks’ vectors

Score vector \( S: \ b_{1,2} \ b_{2,3} \ b_{3,4} \ \cdots \)
The TextTiling algorithm (Hearst 1994, 1997)

Score vector $S$:  $b_{1,2}  \ b_{2,3}  \ b_{3,4}  \  \cdots$

1. Smooth $S$ using average smoothing over window size $a$ to get $\hat{S}$.
2. Set number of boundaries $B$ as $\mu(\hat{S}) - \frac{\sigma(\hat{S})}{2}$
3. Score each boundary $b_i$ using $(b_{i-1} - b_i) + (b_{i+1} - b_i)$
4. Choose the top $B$ boundaries by these scores.
**Dotplotting (Reynar 1994, 1998)**

bulldogs  bulldogs  fight  also  fight | buffalo  that  buffalo  buffalo  also  buffalo

1     2     3     4     5          6     7     8     9     10     11

Where word $w$ appears in positions $x$ and $y$ in a single document, add points $(x, x), (y, y), (x, y)$, and $(y, x)$:
Dotplotting (Reynar 1994, 1998)

<table>
<thead>
<tr>
<th>bulldogs bulldogs fight also fight</th>
<th>buffalo that buffalo buffalo also buffalo</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5</td>
<td>6 7 8 9 10 11</td>
</tr>
</tbody>
</table>

![Dotplot Diagram]

Figure 1: The dotplot of four concatenated *Wall Street Journal* articles. ◯ = actual doc. boundary
Dotplotting (Reynar 1994, 1998)

<table>
<thead>
<tr>
<th>bulldogs</th>
<th>bulldogs</th>
<th>fight</th>
<th>also</th>
<th>fight</th>
<th>buffalo</th>
<th>that</th>
<th>buffalo</th>
<th>buffalo</th>
<th>also</th>
<th>buffalo</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
</tbody>
</table>

Definition (Minimize the density of the regions around the sentences)

- \( n \) = the length of the concatenated texts
- \( m \) = the vocabulary size
- \( \text{Boundaries} \) initialized as \([0]\)
- \( P_j = \text{Boundaries} + j \)
- Vector of length \( m \) containing the number of times each vocab item occurs between positions \( x \) and \( y \)

For a desired number of boundaries \( B \), use dynamic programming to find the \( B \) indices that minimize

\[
\sum_{j=2}^{\lfloor P \rfloor} \frac{V_{P_{j-1}, P_j} \cdot V_{P_j, n}}{(P_j - P_{j-1})(n - P_j)}
\]

Examples (Vocab = (also, buffalo, bulldogs, fight, that))

- \( P = [0, 5] \Rightarrow \frac{[1, 0, 2, 2, 0] \cdot [1, 4, 0, 0, 1]}{(5 - 0)(11 - 5)} = 0.03 \)
- \( P = [0, 6] \Rightarrow \frac{[1, 1, 2, 2, 0] \cdot [1, 3, 0, 0, 1]}{(6 - 0)(11 - 6)} = 0.13 \)
Divisive clustering (Choi 2000)

1. Compare all sentences pairwise for cosine similarity, to create a matrix of similarity values.

   For each value $s$, find the $n \times n$ submatrix $N_s$ with $s$ at its center and replace $s$ with the value

   $$\frac{|\{s' \in N_s : s' < s\}|}{n^2}$$

2. Apply something akin to Reynar’s algorithm to find the cluster boundaries (which are clearer as a result of the local smoothing)

Choi (2000) reports substantial accuracy gains over both TextTiling and dotplotting.
Supervised

1. Label segment boundaries in training and test set.
2. Extract features in training: generally a superset of the features used by unsupervised approaches.
3. Fit a classifier model (NaiveBayes, MaxEnt, SVM, …).
4. In testing, apply feature to predict boundaries.

(Manning 1998; Beeferman et al. 1999; Sharp and Chibelushi 2008)

(Slide from Dan Jurafsky.)
Evaluation: WindowDiff (Pevzner and Hearst 2002)

**Definition (WindowDiff)**

- \( b(i, j) = \) the number of boundaries between text positions \( i \) and \( j \)
- \( N = \) the number of sentences

\[
\text{WindowDiff}(\text{ref}, \text{hyp}) = \frac{1}{N - k} \sum_{i=1}^{N-k} \left( b(\text{ref}_i, \text{ref}_{i+k}) - b(\text{hyp}_i, \text{hyp}_{i+k}) \right) \\
\]

Return values: 0 = all labels correct; 1 = no labels correct

---

**Figure 21.2**  The WindowDiff algorithm, showing the moving window sliding over the hypothesis string, and the computation of \(|r_i - h_i|\) at four positions. After Pevzner and Hearst (2002).

(Jurafsky and Martin 2009:§21)
Discourse coherence theories

- Halliday and Hasan (1976): Additive, Temporal, Causal, Adversative
- Martin (1992): Addition, Temporal, Consequential, Comparison
- Kehler (2002): Result, Explanation, Violated Expectation, Denial of Preventer, Parallel, Contrast (i), Contrast (ii), Exemplification, Generalization, Exception (i), Exception (ii), Elaboration, Occasion (i), Occasion (ii)
- Hobbs (1985): Occasion, Cause, Explanation, Evaluation Background, Exemplification, Elaboration, Parallel, Contrast, Violated Expectation
- Wolf and Gibson (2005): Condition, Violated expectation, Similarity, Contrast, Elaboration, Example, Elaboration, Generalization, Attribution, Temporal Sequence, Same
Rhetorical Structure Theory (RST)

Relations hold between adjacent spans of text: the nucleus and the satellite. Each relation has five fields: constraints on nucleus, constraints on satellite, constraints on nucleus–satellite combination, effect, and locus of effect.

Table 1. *Organization of the relation definitions*

<table>
<thead>
<tr>
<th>Circumstance</th>
<th>Antithesis and Concession</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solutionhood</td>
<td>Antithesis</td>
</tr>
<tr>
<td>Elaboration</td>
<td>Concession</td>
</tr>
<tr>
<td>Background</td>
<td>Condition and Otherwise</td>
</tr>
<tr>
<td>Enablement and Motivation</td>
<td>Condition</td>
</tr>
<tr>
<td>Enablement</td>
<td>Otherwise</td>
</tr>
<tr>
<td>Motivation</td>
<td>Interpretation and Evaluation</td>
</tr>
<tr>
<td>Evidence and Justify</td>
<td>Interpretation</td>
</tr>
<tr>
<td>Evidence</td>
<td>Evaluation</td>
</tr>
<tr>
<td>Justify</td>
<td>Restatement and Summary</td>
</tr>
<tr>
<td>Relations of Cause</td>
<td>Restatement</td>
</tr>
<tr>
<td>Volitional Cause</td>
<td>Summary</td>
</tr>
<tr>
<td>Non-Volitional Cause</td>
<td>Other Relations</td>
</tr>
<tr>
<td>Volitional Result</td>
<td>Sequence</td>
</tr>
<tr>
<td>Non-Volitional Result</td>
<td>Contrast</td>
</tr>
<tr>
<td>Purpose</td>
<td></td>
</tr>
</tbody>
</table>

(Mann and Thompson 1988)
Coherence structures

From Wolf and Gibson (2005)

1. Mr. Baker’s assistant for inter-American affairs,
2. Bernard Aronson
3. while maintaining
4. that the Sandinistas had also broken the cease-fire,
5. acknowledged:

“It’s never very clear who starts what.”

Figure 5
Coherence graph for example (23) with discourse segment 1 split into two segments. expv = violated expectation; elab = elaboration; attr = attribution.
Features for coherence recognition (complete in class)

- Addition
- Temporal
- Contrast
- Causation

Parallel trees identify events:
- ARG
- R nears + contrast
- ROLE CMP
- WordNet rels
- Transition words
- Punctuation
- Times (RETS / verbal tense)
The Penn Discourse Treebank 2.0 (Webber et al. 2003)

- Large-scale effort to identify the coherence relations that hold between pieces of information in discourse.
- Available from the Linguistic Data Consortium.
- Annotators identified spans of text as the coherence relations. Where the relation was implicit, they picked their own lexical items to fill the role.

**Example**

\[
[\text{Arg}_1 \text{ that hung over parts of the factory }] \\
\text{even though} \\
[\text{Arg}_2 \text{ exhaust fans ventilated the area }].
\]
A complex example

[Arg₁ Factory orders and construction outlays were largely flat in December ]

while

purchasing agents said

[Arg₂ manufacturing shrank further in October ].

Explicit

Factory orders and construction outlays were largely flat in September, while purchasing agents said manufacturing shrank further in October

Source: OE/sqj_0078
The overall structure of examples

Don’t try to take it all in at once. It’s too big! Figure out what question you want to address and then focus on the parts of the corpus that matter for it. A brief run-down:

- **Relation-types**: Explicit, Implicit, AltLex, EntRel, NoRel
- **Connective semantics**: hierarchical; lots of levels of granularity to work with, from four abstract classes down to clusters of phrases and lexical items
- **Attribution**: tracking who is committed to what
- **Structure**: Every piece of text is associated with a set of subtrees from the WSJ portion of the Penn Treebank 3.
### Connectives

<table>
<thead>
<tr>
<th>PDTB relation</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explicit</td>
<td>18,459</td>
</tr>
<tr>
<td>Implicit</td>
<td>16,053</td>
</tr>
<tr>
<td>AltLex</td>
<td>624</td>
</tr>
<tr>
<td>EntRel</td>
<td>5,210</td>
</tr>
<tr>
<td>NoRel</td>
<td>254</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>40,600</strong></td>
</tr>
</tbody>
</table>
Explicit connectives

[Arg₁ that hung over parts of the factory ]
even though
[Arg₂ exhaust fans ventilated the area ].

Explicit
that hung over parts of the factory, even though exhaust fans ventilated the area
Source: 00/wsj_0003
Explicit connectives
Implicit connectives

\[ \text{Arg}_1 \text{ Some have raised their cash positions to record levels }. \]
Implicit = BECAUSE
\[ \text{Arg}_2 \text{ High cash positions help buffer a fund when the market falls }. \]

Source: 09/wsj_0083
Implicit connectives

Relation

Implicit

Attribution

i

j

(free text)

(free text)

TEMPORAL
* Asynchronous
* Synchronous
  -- precedence
  -- succession

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

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

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

RAW FILE
the offset of the first character of Arg2 of the Implicit connective
the sentence number of Arg2

PTB 2.0 FILE
the offset of the first character of Arg2 of the Implicit connective
the sentence number of Arg2
AltLex connectives

[Arg₁ Ms. Bartlett's previous work, which earned her an international reputation in the non-horticultural art world, often took gardens as its nominal subject ].

[Arg₂ Mayhap this metaphorical connection made the BPC Fine Arts Committee think she had a literal green thumb ].

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
AltLex connectives

TEMPORAL
- Asynchronous
- Synchronous
  -- precedence
  -- succession

CONTINGENCY
- Cause
  -- reason
  -- result
- Pragmatic Cause
  -- justification
- Condition
  -- hypothetical
  -- general
  -- unreal present
  -- unreal past
  -- factual present
  -- factual past

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

EXPANSION
- Conjunction
- Instantiation
- Restatement
  -- specification
  -- equivalence
  -- generalization
- Alternative
  -- conjunctive
  -- disjunctive
  -- chosen alternative
- Exception
  -- List
Connectives and their semantics

**Figure 1: Hierarchy of sense tags**

(from Prasad et al. 2008)
The relationship between relation-types and connectives

<table>
<thead>
<tr>
<th></th>
<th>Comparison</th>
<th>Contingency</th>
<th>Expansion</th>
<th>Temporal</th>
</tr>
</thead>
<tbody>
<tr>
<td>AltLex</td>
<td>46</td>
<td>275</td>
<td>217</td>
<td>86</td>
</tr>
<tr>
<td>Explicit</td>
<td>5471</td>
<td>3250</td>
<td>6298</td>
<td>3440</td>
</tr>
<tr>
<td>Implicit</td>
<td>2441</td>
<td>4185</td>
<td>8601</td>
<td>826</td>
</tr>
</tbody>
</table>
The distribution of semantic classes
Connectives by relation type

(a) Explicit.

(b) Implicit.

(c) AltLex.

Figure: Wordle representations of the connectives, by relation type.
EntRel and NoRel

[Arg₁] 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.

[Arg₂] Mr. Milgrim succeeds David Berman, who resigned last month.
Arguments
Attributions

[Arg₁ Factory orders and construction outlays were largely flat in December ]
while (Comparison:Contrast:Juxtaposition)
purchasing agents said
[Arg₂ manufacturing shrank further in October ].
Attributions

Attribution strings

researchers said
A Lorillard spokeswoman said
A Lorillard spokeswoman said
said Darrell Phillips, vice president of human resources for Hollingsworth & Vose
said Darrell Phillips, vice president of human resources for Hollingsworth & Vose
Longer maturities are thought
Shorter maturities are considered
considered by some
said Brenda Malizia Negus, editor of Money Fund Report
the Treasury said
The Treasury said
Newsweek said
said Mr. Spoon
According to Audit Bureau of Circulations
According to Audit Bureau of Circulations
saying that

:.
:.
Some informal experimental results: experimental set-up

- Training set of 2,400 examples: 600 randomly chosen examples from each of the four primary PDTB semantic classes: Comparison, Contingency, Expansion, Temporal.
- Test set of 800 examples: 200 randomly chosen examples from each of the four primary semantic classes.
- The students in my LSA class ‘Computational Pragmatics’ formed two teams, and I was a team one one,

and each team specified features, which I implemented using NLTK Python’s MaxEnt interface.
Some informal experimental results: Team Potts

Accuracy: 0.41
Train set accuracy: 1.0

1. **Verb pairs**: features for verb pairs (V1, V2) where V1 was drawn from Arg1 and V2 from Arg2.

2. **Inquirer pairs**: features for the cross product of the Harvard Inquirer semantic classes for Arg1 and Arg2 (after Pitler et al. 2009).
Some informal experimental results: Team Banana Wugs

Accuracy: 0.34
Train set accuracy: 0.37

1. **Negation**: features capturing (sentential and constituent) negation balances and imbalances across the Args.

2. **Sentiment**: A separate sentiment score for each Arg.

3. **Overlap**: the cardinality of the intersection of the Arg1 and Arg2 words divided by their union.

4. **Structural complexity**: features capturing, for each Arg, whether it has an embedded clause, the number of embedded clauses, and the height of its largest tree.

5. **Complexity ratios**: a feature for log of the ratio of the lengths (in words) of the two Args, a feature for the ratio of the clause-counts for the two Args, and a feature for the ratio of the max heights for the two Args.

6. **Pronominal subjects**: a pair-feature capturing whether the subject of the Arg is pronominal (pro) or non-pronominal (non-pro). The features are pairs from \{pro, non-pro\} × \{pro, non-pro\}.

7. **It seems**: returns False if the first argument of the second bigram is not it seems. features

8. **Tense agreement**: a feature for the degree to which the verbal nodes in the two Args have the same tense.

9. **Modals**: a pair-feature capturing whether Arg contains a modal (modal) or not (non-modal). The features are pairs from \{modal, non-modal\} × \{modal, non-modal\}.
Some informal experimental results: Team Banana Slugs

Accuracy: 0.38
Train set accuracy: 0.73

1. **Negation**: for each Arg, a feature for whether it was negated and the number of negation it contains. Also, a feature capturing negation balance/imbalance across the Args.

2. **Main verbs**: for each Arg, a feature for its main-verb. Also, a feature returning True of the two Args’ main verbs match, else False.

3. **Length ratio**: a feature for the ratio of the lengths (in words) of Arg1 and Arg2.

4. **WordNet antonyms**: the number of words in Arg2 that are antonyms of a word in Arg1.

5. **Genre**: a feature for the genre of the file containing the example.

6. **Modals**: for each Arg, the number of modals in it.

7. **WordNet hypernym counts**: for Arg1, a feature for the number of words in Arg2 that are hypernyms of a word in Arg1, and ditto for Arg2.

8. **N-gram features**: for each Arg, a feature for each unigram it contains. (The team suggested going to 2- or 3-grams, but I called a halt at 1 because the data-set is not that big.)
Some informal experimental results: Who won?

Accuracy: 0.41  
Train set accuracy: 1.0

Accurate: 0.34  
Train set accuracy: 0.37

Accuracy: 0.38  
Train set accuracy: 0.73

Feature count: 632,559

Feature count: 116

Feature count: 1,824
Unsupervised discovery of coherence relations (Marcu and Echihabi 2002)

Marcu and Echihabi (2002) focus on four coherence relations that can be informally mapped to coherence relations from other theories:

<table>
<thead>
<tr>
<th>CONTRAST</th>
<th>CAUSE-EXPLANATION-EVIDENCE</th>
<th>ELABORATION</th>
<th>CONDITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANTITHESIS (M&amp;T)</td>
<td>EVIDENCE (M&amp;T)</td>
<td>ELABORATION (M&amp;T)</td>
<td>CONDITION (M&amp;T)</td>
</tr>
<tr>
<td>CONCESSION (M&amp;T)</td>
<td>VOLITIONAL-CAUSE (M&amp;T)</td>
<td>EXPANSION (Ho)</td>
<td></td>
</tr>
<tr>
<td>OTHERWISE (M&amp;T)</td>
<td>NONVOLITIONAL-CAUSE (M&amp;T)</td>
<td>EXEMPLIFICATION (Ho)</td>
<td></td>
</tr>
<tr>
<td>CONTRAST (M&amp;T)</td>
<td>VOLITIONAL-RESULT (M&amp;T)</td>
<td>ELABORATION (A&amp;L)</td>
<td></td>
</tr>
<tr>
<td>VIOLATED EXPECTATION (Ho)</td>
<td>NONVOLITIONAL-RESULT (M&amp;T)</td>
<td>EXPLANATION (A&amp;L)</td>
<td></td>
</tr>
<tr>
<td>( CAUSAL</td>
<td>ADDITIVE ) - ( SEMANTIC</td>
<td>PRAGMATIC ) - NEGATIVE (K&amp;S)</td>
<td>CONTINGENCY: Cause, Pragmatic cause</td>
</tr>
<tr>
<td>Comparison:Contrast</td>
<td>CAUSAL - (SEMANTIC</td>
<td>PRAGMATIC ) - POSITIVE (K&amp;S)</td>
<td>Expansion:Elaboration</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Contingency: Condition, Pragmatic condition</td>
</tr>
</tbody>
</table>

Table 1: Relation definitions as union of definitions proposed by other researchers (M&T – (Mann and Thompson, 1988); Ho – (Hobbs, 1990); A&L – (Lascarides and Asher, 1993); K&S – (Knott and Sanders, 1998)).

Possible PDTB mapping given in red; might want to use to the supercategories.
Automatically collected labels

**Data**

- RAW: 41 million sentences ($\approx$1 billion words) from a variety of LDC corpora
- BLIPP: 1.8 million Charniak parsed sentences

**Labeling method**

1. Extract all sentences matching one of the patterns.
2. Label the connective with the name of the pattern.
3. Treat everything before the connective as Arg1 and everything after it as Arg2.

<table>
<thead>
<tr>
<th>Labeling Type</th>
<th>Number of Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONTRAST — 3,881,588 examples</td>
<td></td>
</tr>
<tr>
<td>[BOS ... EOS] [BOS But ... EOS]</td>
<td></td>
</tr>
<tr>
<td>[BOS ... ] [but ... EOS]</td>
<td></td>
</tr>
<tr>
<td>[BOS ... ] [although ... EOS]</td>
<td></td>
</tr>
<tr>
<td>[BOS Although ... ] [ ... EOS]</td>
<td></td>
</tr>
<tr>
<td>CAUSE-EXPLANATION-EVIDENCE — 889,946 examples</td>
<td></td>
</tr>
<tr>
<td>[BOS ... ] [because ... EOS]</td>
<td></td>
</tr>
<tr>
<td>[BOS Because ... ] [ ... EOS]</td>
<td></td>
</tr>
<tr>
<td>[BOS ... EOS] [BOS Thus, ... EOS]</td>
<td></td>
</tr>
<tr>
<td>CONDITION — 1,203,813 examples</td>
<td></td>
</tr>
<tr>
<td>[BOS If ... ] [ ... EOS]</td>
<td></td>
</tr>
<tr>
<td>[BOS If ... ] [then ... EOS]</td>
<td></td>
</tr>
<tr>
<td>[BOS ... ] [if ... EOS]</td>
<td></td>
</tr>
<tr>
<td>ELABORATION — 1,836,227 examples</td>
<td></td>
</tr>
<tr>
<td>[BOS ... EOS] [BOS ... for example ... EOS]</td>
<td></td>
</tr>
<tr>
<td>[BOS ... ] [which ... ]</td>
<td></td>
</tr>
<tr>
<td>NO-RELATION-SAME-TEXT — 1,000,000 examples</td>
<td></td>
</tr>
<tr>
<td>Randomly extract two sentences that are more than 3 sentences apart in a given text.</td>
<td></td>
</tr>
<tr>
<td>NO-RELATION-DIFFERENT-TEXTS — 1,000,000 examples</td>
<td></td>
</tr>
<tr>
<td>Randomly extract two sentences from two different documents.</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Patterns used to automatically construct a corpus of text span pairs labeled with discourse relations.
Naive Bayes model

1. \( \text{count}(w_i, w_j, r) = \) the number of times that word \( w_i \) occurs in Arg1 and \( w_j \) occurs in Arg2 with coherence relation \( r \).

2. \( W = \) the full vocabulary

3. \( R = \) the set of coherence relations

4. \( N = \sum_{(w_i, w_j) \in W \times W, r \in R} \text{count}(w_i, w_j, r) \)

5. \( P(r) = \frac{\sum_{(w_i, w_j) \in W \times W} \text{count}(w_i, w_j, r)}{N} \)

6. Estimate \( P((w_i, w_j) | r) \) with

\[
\frac{\text{count}(w_i, w_j, r) + 1}{\sum_{(w_x, w_y) \in W \times W} \text{count}(w_x, w_y, r) + N}
\]

7. Maximum likelihood estimates for example with \( W_1 \) the words in Arg1 and \( W_2 \) the words in Arg2:

\[
\arg \max_r \left[ P(r) \prod_{(w_i, w_j) \in W_1 \times W_2} P((w_i, w_j) | r) \right]
\]

(Connectives are excluded from these calculations, since they were used to obtain the labels.)
Results for pairwise classifiers

<table>
<thead>
<tr>
<th></th>
<th>CONTRAST</th>
<th>CEV</th>
<th>COND</th>
<th>ELAB</th>
<th>NO-REL-SAME-TEXT</th>
<th>NO-REL-DIFF-TEXTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONTRAST</td>
<td>-</td>
<td>87</td>
<td>74</td>
<td>82</td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td>CEV</td>
<td>76</td>
<td>93</td>
<td>75</td>
<td>76</td>
<td>74</td>
<td>74</td>
</tr>
<tr>
<td>COND</td>
<td>89</td>
<td>93</td>
<td>69</td>
<td>76</td>
<td>71</td>
<td>71</td>
</tr>
<tr>
<td>ELAB</td>
<td>89</td>
<td>64</td>
<td>76</td>
<td>66</td>
<td>57</td>
<td>57</td>
</tr>
</tbody>
</table>

Table 3: Performances of classifiers trained on the Raw corpus. The baseline in all cases is 50%.

<table>
<thead>
<tr>
<th></th>
<th>CONTRAST</th>
<th>CEV</th>
<th>COND</th>
<th>ELAB</th>
<th>NO-REL-SAME-TEXT</th>
<th>NO-REL-DIFF-TEXTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONTRAST</td>
<td>-</td>
<td>62</td>
<td>58</td>
<td>64</td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td>CEV</td>
<td>69</td>
<td>82</td>
<td>63</td>
<td>78</td>
<td>68</td>
<td>68</td>
</tr>
<tr>
<td>COND</td>
<td>78</td>
<td>63</td>
<td>57</td>
<td>78</td>
<td>57</td>
<td>57</td>
</tr>
<tr>
<td>ELAB</td>
<td>78</td>
<td>64</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
</tr>
</tbody>
</table>

Table 4: Performances of classifiers trained on the BLIPPP corpus. The baseline in all cases is 50%.

Systems trained on the smaller, higher-precision BLIPPP corpus have lower overall accuracy, but they perform better with less data than those trained on the RAW corpus.

Figure 1: Learning curves for the ELABORATION vs. CAUSE-EXPLANATION-EVIDENCE classifiers, trained on the Raw and BLIPPP corpora.
Results for the RST corpus of Carlson et al. 2001

For this experiment, the classifiers were trained on the RAW corpus, with the connectives included as features. Only RST examples involving (approximations of) the four relations used above were in the test set.

<table>
<thead>
<tr>
<th>Relation</th>
<th># Test Cases</th>
<th>CONTR</th>
<th>CEV</th>
<th>COND</th>
<th>ELAB</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONTR</td>
<td>238</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CEV</td>
<td>307</td>
<td>63 56</td>
<td></td>
<td>80 65</td>
<td>64 88</td>
</tr>
<tr>
<td>COND</td>
<td>125</td>
<td>87 71</td>
<td></td>
<td>76 85</td>
<td>87 93</td>
</tr>
<tr>
<td>ELAB</td>
<td>1761</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Performances of Raw-trained classifiers on manually labeled RST relations that hold between elementary discourse units. Performance results are shown in bold; baselines are shown in normal fonts.

**Identifying implicit relations**

The RAW-trained classifier is able to accurately guess a large number of implicit examples, essentially because it saw similar examples with an overt connective (which served as the label).

**In sum**: an example of the ‘unreasonable effectiveness of data’ (Banko and Brill 2001; Halevy et al. 2009).
Data and tools

- **Penn Discourse Treebank 2.0**
  - LDC: [http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2008T05](http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2008T05)
  - Project page: [http://www.seas.upenn.edu/~pdtb/](http://www.seas.upenn.edu/~pdtb/)
  - Python tools/code: [http://compprag.christopherpotts.net/pdtb.html](http://compprag.christopherpotts.net/pdtb.html)

- **Rhetorical Structure Theory**
  - LDC: [http://www.ldc.upenn.edu/Catalog/catalogEntry.jsp?catalogId=LDC2002T07](http://www.ldc.upenn.edu/Catalog/catalogEntry.jsp?catalogId=LDC2002T07)
  - Project page: [http://www.sfu.ca/rst/](http://www.sfu.ca/rst/)
Prospects

Text segmentation

Seems to have fallen out of fashion, but obviously important to many kinds of information extraction — probably awaiting a breakthrough idea.

Discourse coherence

On the rise in linguistics but perhaps not in NLP. Essential to all aspects of NLU, though, so a breakthrough would probably have widespread influence.
References


Halevy, Alon; Peter Norvig; and Fernando Pereira. 2009. The unreasonable effectiveness of data. *IEEE Intelligent Systems* 24(2):8–12.


References II

References III


Prasad, Rashmi; Nikhil Dinesh; Alan Lee; Eleni Miltsakaki; Livio Robaldo; Aravind Joshi; and Bonnie Webber. 2008. The Penn Discourse Treebank 2.0. In Nicoletta Calzolari; Khalid Choukri; Bente Maegaard; Joseph Mariani; Jan Odijk; Stelios Piperidis; and Daniel Tapias, eds., Proceedings of the Sixth International Language Resources and Evaluation (LREC’08). Marrakech, Morocco: European Language Resources Association (ELRA). URL http://www.lrec-conf.org/proceedings/lrec2008/.


References IV