Discourse Coherence

20 April 2021
We’ve been looking at linguistic phenomena at the level of individual words or sentences, but language doesn’t end there.
**Discourse** is any linguistic unit that consists of multiple sentences, e.g.,

*On Monday, John went to Einstein’s. He wanted to buy lunch. But the cafe was closed. That made him angry, so the next day he went to Green Street instead.*
To understand discourse, we need to do coreference resolution, as discussed last time:

On Monday, John went to Einstein’s. He wanted to buy lunch. But the cafe was closed. That made him angry, so the next day he went to Green Street instead.
To understand discourse, we need to do coreference resolution, as discussed last time:

*On Monday, John went to Einstein’s. He wanted to buy lunch. But the cafe was closed. That made him angry, so the next day he went to Green Street instead.*
To understand discourse, we also need to identify discourse ("coherence") relations:

- On Monday, John went to Einstein’s.
- He wanted to buy lunch.

(reason)
Discourse segmentation: chunking texts into coherent units.

Discourse coherence: Characterizing the meaning relationships between clauses in text.
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 Hendrick¹,³, 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, 1600 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 screen-film 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.

Highly structured abstract from a paper on PubMed
Identification of Genes Required for the Function of Non-Race-Specific mlo Resistance to Powdery Mildew in Barley

A. Freialdenhoven, C. Peterhansel, J. Kurth, F. Kreuzaler and P. Schulze-Lefert
Rheinisch-Westfälische Technische Hochschule Aachen, Department of Biology I, Worringen 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/or 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.
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?

Report abuse | Permalink

5-star Amazon review
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 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?

3-star Amazon review
Discourse matters for natural language understanding:

Most information isn’t contained in a single sentence.
The system has to aggregate information across paragraphs or entire documents.
Discourse also matters for natural language generation:

When systems generate text, that text needs to be easy to understand, i.e., coherent.

What makes text coherent?
Coherence examples

Sam brushed his teeth.

He got into bed.

He felt a certain ennui.
Coherence examples

Sam brushed his teeth. [then]

He got into bed. [then]

He felt a certain ennui.
Coherence examples

*Sue was feeling ill.*

*She decided to stay home from work.*
Coherence examples

Sue was feeling ill. [so]

She decided to stay home from work.
Coherence examples

*Sue likes bananas.*

*Jill does not.*
Coherence examples

Sue likes bananas. [but]

Jill does not.
Coherence examples

The senator introduced a new initiative.

He hoped to please undecided voters.
Coherence examples

The senator introduced a new initiative. [because]

He hoped to please undecided voters.
Coherence examples

*Linguists like quantifiers.*

*In his lectures, Richard talked only about “every” and “most”.*
Coherence examples

Linguists like quantifiers. [for example]

In his lectures, Richard talked only about “every” and “most”.

Coherence examples

In his lectures, Richard talked only about “every” and “most”.

Linguists like quantifiers.
Coherence examples

In his lectures, Richard talked only about “every” and “most”. [in general]

Linguists like quantifiers.
Today we’ll look at

Unsupervised and supervised discourse segmentation
Discourse coherence theories
The Penn Discourse Treebank data set
Discourse segmentation
Discourse segmentation is the task of separating a document into a linear sequence of subtopics.

This is simpler than detecting embedded structures.
Stargazer

Intro – the search for life in space

The moon’s chemical composition

How early earth-moon proximity shaped the moon

How the moon helped life evolve on earth

Improbability of the earth-moon system

Binary/trinary star systems make life unlikely

The low probability of nonbinary/trinary systems

Properties of earth’s sun that facilitate life

Summary

Figure 5
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.
Intuition of cohesion-based segmentation:

Sentences or paragraphs in a subtopic are cohesive with each other, but not with paragraphs in a neighboring subtopic.

Procedure:

Measure cohesion between neighboring sentences.
Segment where there's a “dip” in cohesion.
The TextTiling algorithm (Hearst 1997)

Score this boundary via cosine similarity between the blocks’ vectors

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

Score this boundary via cosine similarity between the blocks’ vectors

<table>
<thead>
<tr>
<th></th>
<th>$w_1$</th>
<th>$w_2$</th>
<th>$w_3$</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_1$</td>
<td>[s_1]</td>
<td>[s_1]</td>
<td>[s_1]</td>
<td>...</td>
</tr>
<tr>
<td>$s_2$</td>
<td>[s_2]</td>
<td>[s_2]</td>
<td>[s_2]</td>
<td>...</td>
</tr>
<tr>
<td>$s_3$</td>
<td>[s_3]</td>
<td>[s_3]</td>
<td>[s_3]</td>
<td>...</td>
</tr>
<tr>
<td>$s_4$</td>
<td>[\vdots]</td>
<td>[\vdots]</td>
<td>[\vdots]</td>
<td>...</td>
</tr>
<tr>
<td>$s_5$</td>
<td>[s_8]</td>
<td>[s_8]</td>
<td>[s_8]</td>
<td>...</td>
</tr>
<tr>
<td>$s_6$</td>
<td>[\vdots]</td>
<td>[\vdots]</td>
<td>[\vdots]</td>
<td>...</td>
</tr>
<tr>
<td>$s_7$</td>
<td>[\vdots]</td>
<td>[\vdots]</td>
<td>[\vdots]</td>
<td>...</td>
</tr>
<tr>
<td>$s_8$</td>
<td>[\vdots]</td>
<td>[\vdots]</td>
<td>[\vdots]</td>
<td>...</td>
</tr>
</tbody>
</table>

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

Score this boundary via cosine similarity between the blocks’ vectors

<table>
<thead>
<tr>
<th></th>
<th>$w_1$</th>
<th>$w_2$</th>
<th>$w_3$</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_1$</td>
<td>$s_1$</td>
<td>$s_1$</td>
<td>$s_1$</td>
<td>...</td>
</tr>
<tr>
<td>$s_2$</td>
<td>$s_2$</td>
<td>$s_2$</td>
<td>$s_2$</td>
<td>...</td>
</tr>
<tr>
<td>$s_3$</td>
<td>$s_3$</td>
<td>$s_3$</td>
<td>$s_3$</td>
<td>...</td>
</tr>
<tr>
<td>$s_4$</td>
<td>$s_4$</td>
<td>$s_4$</td>
<td>$s_4$</td>
<td>...</td>
</tr>
<tr>
<td>$s_5$</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$s_6$</td>
<td>$s_9$</td>
<td>$s_9$</td>
<td>$s_9$</td>
<td>...</td>
</tr>
<tr>
<td>$s_7$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_8$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_9$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Score vector $S$: $b_{1,2}$ $b_{2,3}$ $b_{3,4}$ ...
The TextTiling algorithm (Hearst 1997)

<table>
<thead>
<tr>
<th>( s_1 )</th>
<th>( s_2 )</th>
<th>( s_3 )</th>
<th>( \cdots )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{sum} )</td>
<td>( s_1 )</td>
<td>( s_2 )</td>
<td>( s_3 )</td>
</tr>
<tr>
<td>( \text{sum} )</td>
<td>( s_2 )</td>
<td>( s_3 )</td>
<td>( s_3 )</td>
</tr>
<tr>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
</tr>
<tr>
<td>( \text{sum} )</td>
<td>( s_{9} )</td>
<td>( s_{9} )</td>
<td>( s_{9} )</td>
</tr>
</tbody>
</table>

Score this boundary via cosine similarity between the blocks’ vectors

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

1. Smooth \( S \) using average smoothing over window size \( a \) to get \( \hat{S} \).
2. Set number of boundaries \( B \) as \( \mu(\hat{S}) - \sigma(\hat{S})/2 \)
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.
The TextTiling algorithm (Hearst 1997)

dot-product computation of similarity between sentences
Source code for nltk.tokenize.texttiling

```python
# Natural Language Toolkit: TextTiling
#
# Copyright (C) 2001-2021 NLTK Project
# Author: George Choudakis
# URL: <http://nltk.org/>
# For license information, see LICENSE.TXT

import re
import math

try:
    import numpy
except ImportError:
    pass

from nltk.tokenize.api import TokenizerI
BLACK_COMPARISON, VOCABULARY_INTRODUCTION = 0, 1
LC, WC = 0, 1
DEFAULT_SMOOTHING = {0}

class TextTilingTokenizer(TokenizerI):
    """Tokenize a document into topical sections using the TextTiling algorithm.
    This algorithm detects subtopic shifts based on the analysis of lexical
    co-occurrence patterns.

    The process starts by tokenizing the text into pseudosentences of
    a fixed size v. Then, depending on the method used, similarity
    scores are assigned at sentence gaps. The algorithm proceeds by
    detecting the peak differences between these scores and marking
    them as boundaries. The boundaries are normalized to the closest
    paragraph break and the segmented text is returned."""
```
Supervised discourse segmentation

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 (naïve Bayes, maximum entropy, etc.)

4. In testing, apply feature to predict boundaries.

Manning 1998;
Beeferman et al. 1999;
Sharp and Chibelushi 2008
Supervised discourse segmentation

Learned features include discourse markers or cue words, e.g.,

Broadcast news
  *Good evening, I’m…*
  *Coming up….*

Science articles
  *First, …*
  *The next topic…*
Evaluation: WindowDiff (Pevzner & Hearst, 2002)

Let $b(i, j)$ be the number of boundaries between text positions $i$ and $j$.

Let $N$ be the number of sentences.

$$\text{WindowDiff}(\text{ref}, \text{hyp}) = \frac{1}{N-k} \sum_{i=1}^{N-k} \left| b(\text{ref}_i, \text{ref}_{i+k}) - b(\text{hyp}_i, \text{hyp}_{i+k}) \right| \neq 0$$

Return values:

0 = all labels correct
1 = no labels correct
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>
<thead>
<tr>
<th>Table 1. Organization of the relation definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circumstance</td>
</tr>
<tr>
<td>Solutionhood</td>
</tr>
<tr>
<td>Elaboration</td>
</tr>
<tr>
<td>Background</td>
</tr>
<tr>
<td>Enablement and Motivation</td>
</tr>
<tr>
<td>Enablement</td>
</tr>
<tr>
<td>Motivation</td>
</tr>
<tr>
<td>Evidence and Justify</td>
</tr>
<tr>
<td>Evidence</td>
</tr>
<tr>
<td>Justify</td>
</tr>
<tr>
<td>Relations of Cause</td>
</tr>
<tr>
<td>Volitional Cause</td>
</tr>
<tr>
<td>Non-Volitional Cause</td>
</tr>
<tr>
<td>Volitional Result</td>
</tr>
<tr>
<td>Non-Volitional Result</td>
</tr>
<tr>
<td>Purpose</td>
</tr>
</tbody>
</table>

*Mann & Thompson, 1988*
Coherence structures

1. a. Mr Baker’s assistant for inter-American affairs,
   b. Bernard Aronson
2. while maintaining
3. that the Sandinistas had also broken the cease-fire,
4. acknowledged:
5. “It’s never very clear who starts what.”

Wolf & Gibson, 2005
Penn Discourse Treebank
The Penn Discourse Treebank 2.0 (Webber et al., 2003) is a large-scale effort to identify the coherence relations that hold between pieces of information in discourse.

Available from the Linguistic Data Consortium (LDC).

Annotators identified spans of text as the coherence relations.

Where the relation was implicit, they picked their own lexical items to fill the role.

An example:

\[
\text{[Arg1 } that \text{ hung over parts of the factory ]}
\]

\text{even though}

\[
\text{[Arg2 exhaust fans ventilated the area ]}.
\]
A complex example

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

while

purchasing agents said

[Arg2 manufacturing shrank further in October].
Factory orders and construction outlays were largely flat in September.

(SBAR (NONE-0) (S (NP-SBJ (NN manufacturing)) (VP (VBD shrank) (ADVP (RB further)) (PP-TMP (IN in) (NP (NNP October))))))

Manufacturing shrank further in October.

(purchasing agents said)

(NP-SBJ (VBG purchasing) (NNS agents))

(VBD said)

Or
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

\[ \text{Arg}_1 \text{ that hung over parts of the factory } \]

\textit{even though}

\[ \text{Arg}_2 \text{ exhaust fans ventilated the area } \].
Implicit connectives

[Arg1 Some have raised their cash positions to record levels ].
Implicit = BECAUSE
[Arg2 High cash positions help buffer a fund when the market falls ].
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 ].
Connectives and their semantics

Prasad et al., 2008
Connectives by relation type

(a) Explicit.

(b) Implicit.

(c) AltLex.
EntRel and NoRel

$[\text{Arg}_1]$ Hale Milgrim, 41 years old, senior vice president, marketing at Elektra Entertainment Inc., was named president of Capitol Records Inc., a unit of this entertainment concern.

$[\text{Arg}_2]$ Mr. Milgrim succeeds David Berman, who resigned last month.
Attributions

\[\text{Arg}_1\text{ Factory orders and construction outlays were largely flat in December }\]

\textit{while} (Comparison:Contrast:Juxtaposition)

\textit{purchasing agents said}

\[\text{Arg}_2\text{ manufacturing shrunk further in October }\].
Attributions

Attribution strings

researchers said
A Lorillard spokeswoman said
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
Newsweek said
said Mr. Spoon
According to Audit Bureau of Circulations
saying that
Conclusion
Data and tools

Penn Discourse Treebank 2.0

   Project page: seas.upenn.edu/~pdtb
   LDC: ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2008T05
   Python tools/code: compprag.christopherpotts.net/pdtb.html

Rhetorical structure theory

   Project page: sfu.ca/rst
   LDC: ldc.upenn.edu/Catalog/catalogEntry.jsp?catalogId=LDC2002T07
Prospects

Text segmentation seems to have fallen out of fashion, but it’s obviously important to many kinds of information extraction – probably awaiting a breakthrough idea.

Discourse coherence is on the rise in linguistics, but perhaps not in NLP. It’s essential to all aspects of NLU, though, so a breakthrough would probably have widespread influence.
Acknowledgments

The lecture incorporates material from:

Daniel Jurafsky and James Martin, Speech and Language Processing
Julia Hockenmaier, University of Illinois
Nancy Ide, Vassar College
Chris Potts, Stanford University