Models of Discourse Structure

Jacob Eisenstein

Georgia Institute of Technology

July 24, 2015

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

Running example #1

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

Translations from French subtitles

Original	**From French**
You clearly don't know who you're talking to,	You do not know to whom you address.
so let me clue you in.	Let me be clear.
I am not in danger, Skyler.	I'm not in danger, Skyler.
I am the danger.	The danger is me.
A guy opens his door and gets shot,	A guy gets off at his door.
and you think that of me?	You see me in his place?
No. I am the one who knocks!	No. It is I who knocks on the door.

Running example #2

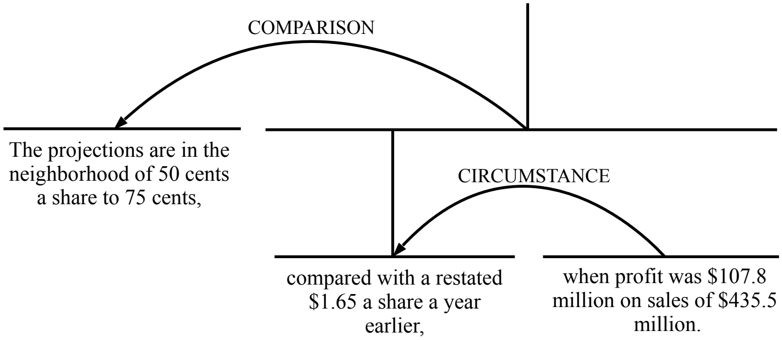
The more people you love, the weaker you are. You'll do things for them that you know you shouldn't do. You'll act the fool to make them happy, to keep them safe. Love no one but your children. On that front, a mother has no choice.

`https://www.youtube.com/watch?v=49_cPvbNA54#t=3m47s`

Translations from French subtitles

Original	From French
The more people you love, the weaker you are.	The more you love people, more you are weak.
You'll do things for them that you know you shouldn't do.	You will do things knowing that you should not do them.
You'll act the fool to make them happy, to keep them safe.	You will play the madness to make them happy, for their protection.
Love no one but your children.	Love nothing more than your children.
On that front, a mother has no choice.	In this world, women have no other choice.

Table of Contents



Bill Mann and Sandra Thompson (1988), **Rhetorical Structure Theory: Towards a functional theory of text organization.**

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.

RST segmentation: example

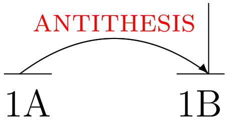
- You clearly don't know who you're talking to,
- so let me clue you in.
- I am not in danger, Skyler.
- I am the danger.
- A guy opens his door
- **and gets shot**,
- **and you think that of me?**
- No.
- I am the one who knocks!

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**.



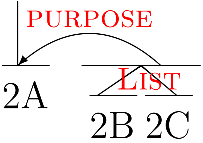
[I am not in danger, Skyler.]^{1A} [I am the danger.]^{1B}

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" ≈ projectivity.

Multinuclear relations

Discourse relations can involve multiple nucleii.



[You'll act the fool]^{2A} [to keep them happy,]^{2B} [to keep them safe.]^{2C}

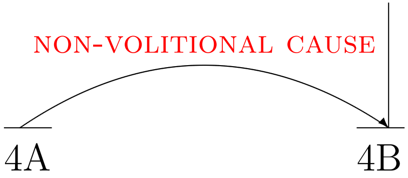
- "A multinuclear relation contains two or more units or spans of equal importance in the discourse."
- "Nuclearity assignment is often determined simultaneously with the rhetorical relation. What counts as a nucleus... can rarely be determined in isolation."

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.

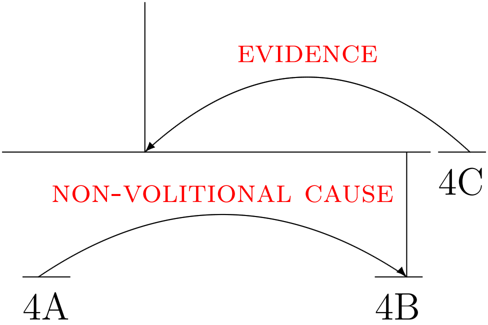
Non-volitional cause



[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.

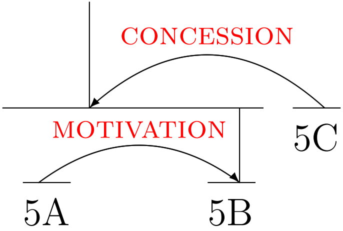
Evidence



[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"

Motivation and concession



[The more people you love, the weaker you are. You'll act the fool to make them happy, to keep them safe.]^{5A} [Love no one but your children.]^{5B} [On that front, a mother has no choice.]^{5C}

- Motivation: "Comprehending S increases R's desire to perform action in N"
- Concession: "W acknowledges a potential or apparent incompatibility between N and S."

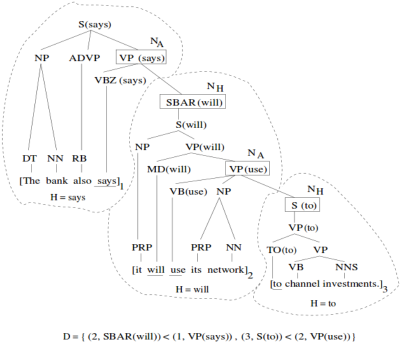
Inter-annotator agreement

Span	Nuclearity	Relation
88.7	77.7	65.8

- **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.

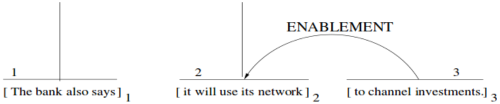
EDU segmentation

Generative model of Soricut & Marcu (2003):



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

Generative RST parsing

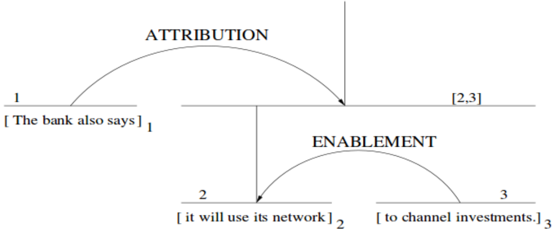


$S1 = 1$

P_s ([2,2,3] | (2,SBAR)<(1,VP), (3,S)<(2,VP)) = 0.47

P_r (ENABLEMENT–NS | S(to)<VP(use)) = 0.88

$S2 = P_s * P_r = 0.40$



P_s([1,1,3] | (2, SBAR) < (1, VP), (3, S) < (2, VP)) = 0.37

P_r (ATTRIBUTION–SN | SBAR(will) < VP(says)) = 0.009

$\text{Score1} = S1*S2* P_s * P_r = 0.001$

Modern RST parsing

Approaches:

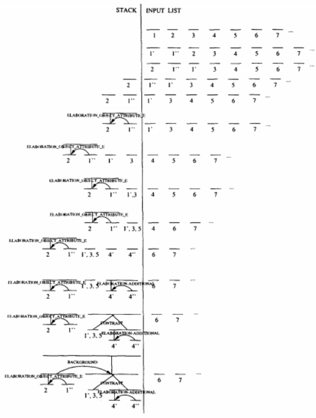
- Classification + CKY (Hernault et al., 2010; Feng & Hirst, 2012)
- Sequence labeling (Ghosh et al., 2012; Joty et al., 2013)
- Shift-reduce (Sagae, 2009; Ji & Eisenstein, 2014; Heilman & Sagae, 2015)
- Representation learning (Ji & Eisenstein, 2014; Li et al., 2014)

Best results (on gold EDU segments)

	Span	Nuclearity	Relation
Feng & Hirst (2014)	**85.7**	71.0	58.2
Li et al. (2014)	82.9	**73.0**	60.6
Ji & Eisenstein (2014)	81.6	71.0	**61.8**

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)



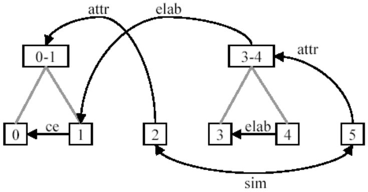
Annotated corpora

- **RST Treebank**: 385 English newswire documents
- **RST Spanish Treebank**: several hundred documents, apparently academic abstracts, `http://corpus.iingen.unam.mx/rst/corpus_en.html`.
- **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.

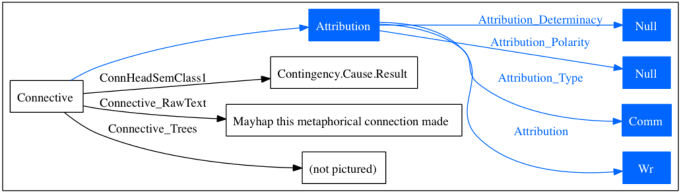
0. Farm prices in October edged up 0.7% from September
1. as raw milk prices continued to rise,
2. the Agriculture Department said.
3. Milk sold to the nations dairy plants and dealers averaged $14.50 for each hundred pounds,
4. up 50 percent from September and up $1.50 from October 1988,
5. the department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



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

Table of Contents

The Penn Discourse Treebank



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

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber (2008). **The Penn Discourse Treebank 2.0**. Proceedings of LREC.

D-LTAG: Lexically anchored discourse relations

- Choosing from the 78 RST relations is hard!
- Let's just assume relations ≈ 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.
- Formally: lexicalized tree-adjoining grammar for discourse, D-LTAG Webber (2004).

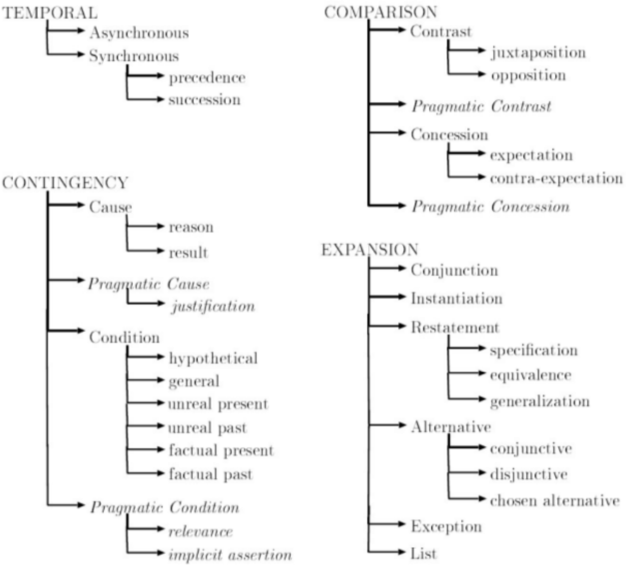
From parsing to chunking

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

> *[PDTB] makes no commitment to any kind of higher-level discourse structure over the discourse relations annotated between individual text spans... PDTB invites experimentation with approaches to high-level topic and functional structuring or to hierarchical structuring...*
>
> *(Prasad et al., 2014)*

PDTB parsing is sometimes called "shallow discourse parsing" (Prasad et al., 2010).

Relation taxonomy



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, **AltLex**;
- or that the sentences are related by shared entities, **EntRel**;
- or that there is no relation, **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?
- <u>No,</u> 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.Instatiation) 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.
- <u>On that front</u> (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):

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

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).
- Determining the argument boundaries is pretty hard: Lin et al. (2010) use classifiers, Wang et al. (2010) use tree kernels.

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.

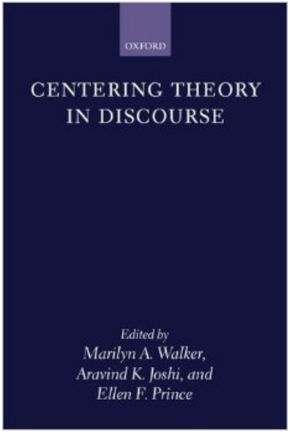
(Evaluation in this area is kind of a mess.)

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



Barbara Grosz, Aravind Joshi, and Scott Weinstein (1995), **Centering: A Framework for Modeling the Local Coherence of Discourse**. Computational Linguistics 21 (2).

- 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.

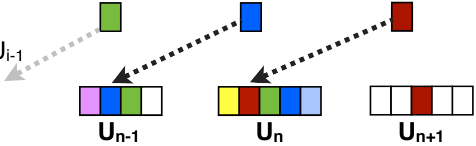
We maintain a list of discourse entities, ranked by salience.
The position in this list determines how easy it is to refer back to an entity in the next utterance.
Each utterance updates this list.

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 $\mathbf{U_n}$ is a partially ordered **list of the entities** mentioned in $\mathbf{U_n}$.

The ordering reflects **salience** within $\mathbf{U_n}$:

subject > direct object > object,….

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

Center realization and pronouns

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

Constraint/Rule 1:

If ***any*** **element of FW($\mathbf{U_{n-1}}$)** is realized as **a pronoun** in U_n,
then the **BW($\mathbf{U_n}$)** 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:

$BW(U_n) = BW(U_{n-1})$. $BW(U_n)$ is highest ranked element in $FW(U_n)$

Sue *gave* ***Joe a dog.***

She *told* ***him*** *to feed* ***it*** *well.* **BW**=Sue, **FW**={Sue, Joe, dog}

She *asked* ***him*** *whether he liked the gift.* **BW**=Sue, **FW**={Sue, Joe, gift}

Center retaining:

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

Sue *gave* ***Joe a dog****.*

She *told* ***him*** *to feed* ***it*** *well.* **BW**=Sue, **FW**={Sue, Joe, dog}

John *asked* ***her*** *what to feed him.* **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.* **BW**=Sue, **FW**={Sue, Joe,dog}

The dog *was very cute.* **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

U_n	C_f	C_b
You don't know who you're talking to,	Skyler, Walter	-
so let me clue you in.	Skyler, Walter	Skyler
I am not in danger, Skyler.	Walter, danger, Skyler	Skyler
I am the danger.	Walter, danger	Walter
A guy opens his door and gets shot,	a guy, the door	-
and you think that of me?	Skyler, Walter	-
No. I am the one who knocks!	Walter	Walter

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.

Entity grid example

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

Entity grid translated from French subtitles

	Skyler	Walter	danger	a guy	the door
You do not know to whom you address.	S	- (O?)	-	-	-
Let me be clear.	- (S?)	O	-	-	-
I am not in danger, Skyler.	X	S	X	-	-
The danger is me.	-	O	S	-	-
A guy gets off at his door,	-	-	-	S	O
you see me in his place?	S	O	-	-	-
No! It is I who knocks on the door	-	S	-	-	X

Entity grid example 2

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

Entity grid translated from French subtitles

	you	people you love
The more you love people, more you are weak.	S	O
You will do things knowing that you should not do them.	S	-
You will play the madness to make them happy, for their protection.	S	O
Love nothing more than your children.	X	-
In this world, women have no other choice	-	-

(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)

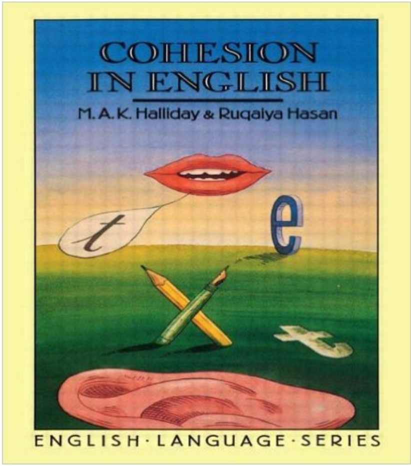
1. A man walked in. He ordered a beer.
2. Every man walked in. ?? He ordered a beer.
3. John can open Bill's safe. He's going to have to get the combination changed soon.[1]

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

[1] L&A attribute this example to Hobbs (1985), which I can't find.

Table of Contents

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]*

[2] Thanks to Joshi, Prasad, and Webber (2006) for reminding me of these quotes.

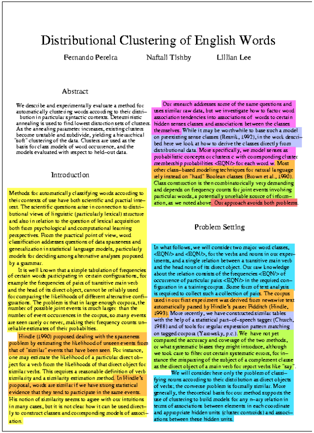
Textiling

Hearst (1997) showed that lexical cohesion could be quantified and used to support unsupervised topic topic segmentation.

```
-------------------------------------------------------------------------------------------------------------
Sentence:        05   10   15   20   25   30   35   40   45   50   55   60   65   70   75   80   85   90   95
-------------------------------------------------------------------------------------------------------------
14       form    1     111 1   1                               1 1     1   1       1     1     1    1
 8  scientist                      11            1   1               1          1      1  1
 5      space 11    1     1                                                            1
25       star    1             1                                       11 22  111112  1 1  1   11 1111     1
 5     binary                                                            11  1          1                    1
 4    trinary                                                             1   1         1                    1
 8 astronomer 1                1                                        1 1           1   1    1  1
 7      orbit   1                    1                                     12    1 1
 6       pull                          2     1 1                               1  1
16     planet   1    1       11               1           1                21  11111                  1     1
 7     galaxy    1                                                  1                1  11     1           1
 4      lunar            1  1     1       1
19       life 1  1  1                              1    11 1  11  1      1                 1 1    1 111  1 1
27       moon        13  1111   1 1 22 21  21    21           11 1
 3       move                                      1   1   1
 7  continent                                      2 1 1 2 1
 3  shoreline                                            12
 6       time                      1               1  1  1     1                                          1
 3      water                                  11           1
 6        say                                 1 1        1        11               1
 3    species                                        1  1  1
-------------------------------------------------------------------------------------------------------------
Sentence:        05   10   15   20   25   30   35   40   45   50   55   60   65   70   75   80   85   90   95
-------------------------------------------------------------------------------------------------------------
```

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)



Distributional Clustering of English Words

Fernando Pereira Naftali Tishby Lillian Lee

Abstract

We describe and experimentally evaluate a method for automatically clustering words according to their distribution in particular syntactic contexts. Deterministic annealing is used to find lowest distortion sets of clusters. As the annealing parameter increases, existing clusters become unstable and subdivide, yielding a hierarchical "soft" clustering of the data. Clusters are used as the basis for class models of word occurrence, and the models evaluated with respect to held–out data.

Introduction

Methods for automatically classifying words according to their contexts of use have both scientific and practial interest. The scientific questions arise in connection to distributional views of linguistic (particularly lexical) structure and also in relation to the question of lexical acquisition both from psychological and computational learning perspectives. From the practical point of view, word classification addresses questions of data sparseness and generalization in statistical language models, particularly models for deciding among alternative analyses proposed by a grammar.

It is well known that a simple tabulation of frequencies of certain words participating in certain configurations, for example the frequencies of pairs of transitive main verb and the head of its direct object, cannot be reliably used for comparing the likelihoods of different alternative configurations. The problem is that in large enough corpora, the number of possible joint events is much larger than the number of event occurrences in the corpus, so many events are seen rarely or never, making their frequency counts unreliable estimates of their probabilties.

Hindle (1990) proposed dealing with the sparseness problem by estimating the likelihood of unseen events from that of "similar" events that have been seen. For instance, one may estimate the likelihood of a particular direct object for a verb from the likelihoods of that direct object for similar verbs. This requires a reasonable definition of verb similarity and a similarity estimation method. In Hindle's proposal, words are similar if we have strong statistical evidence that they tend to participate in the same events. His notion of similarity seems to agree with our intuitions in many cases, but it is not clear how it can be used directly to construct classes and corresponding models of association.

Our research addresses some of the same questions and uses similar raw data, but we investigate how to factor word association tendencies into associations of words to certain hidden senses classes and associations between the classes themselves. While it may be worthwhile to base such a model on preexisting sense classes (Resnik, 1992), in the work described here we look at how to derive the classes directly from distributional data. More specifically, we model senses as probabilistic concepts or clusters c with corresponding cluster membership probabilities <EQN/> for each word w. Most other class–based modeling techniques for natural language rely instead on "hard" Boolean classes (Brown et al., 1990). Class constuction is then combinatorically very demanding and depends on frequency counts for joint events involving particular words, a potentially unreliable source of information, as we noted above. Our approach avoids both problems.

Problem Setting

In what follows, we will consider two major word classes, <EQN/> and <EQN/>, for the verbs and nouns in our experiments, and a single relation between a transitive main verb and the head noun of its direct object. Our raw knowledge about the relation consists of the frequencies <EQN/> of occurrence of particular pairs <EQN/> in the required configuration in a training corpus. Some form of text analysis is required to collect such a collection of pairs. The corpus used in our first experiment was derived from newswire text automatically parsed by Hindle's parser Fidditch (Hindle, 1993). More recently, we have constructed similar tables with the help of a statistical part–of–speech tagger (Church, 1988) and of tools for regular expression pattern matching on tagged corpora (Yarowsky, p.c.). We have not yet compared the accuracy and coverage of the two methods, or what systematic biases they might introduce, although we took care to filter out certain systematic errors, for instance the misparsing of the subject of a complement clause as the direct object of a main verb for report verbs like "say".

We will consider here only the problem of classifying nouns according to their distribution as direct objects of verbs; the converse problem is formally similar. More generally, the theoretical basis for our method supports the use of clustering to build models for any n–ary relation in terms of associations between elements in each coordinate and appropriate hidden units (cluster centroids) and associations between these hidden units.

(Teufel, 1999)

Functional discourse structure

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 improve information extraction in conventionalized domains:

- biomedical abstracts (McKnight & Srinivasan, 2003)
- legal documents (Palau & Moens, 2009)

Conventionalized topic structures

	Wisconsin	Louisiana	Vermont
1	Etymology	Etymology	Geography
2	History	Geography	History
3	Geography	History	Demographics
4	Demographics	Demographics	Economy
5	Law and government	Economy	Transportation
6	Economy	Law and government	Media
7	Municipalities	Education	Utilities
8	...	...	...

Wikipedia articles about US states

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)

Summary

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 I

Al-Saif, A. & Markert, K. (2010). The leeds arabic discourse treebank: Annotating discourse connectives for arabic. In *LREC*.

Barzilay, R. & Lapata, M. (2008). Modeling local coherence: An Entity-Based approach. *Computational Linguistics*, *34*(1), 1–34.

Ben, G., Xiong, D., Teng, Z., Lü, Y., & Liu, Q. (2013). Bilingual lexical cohesion trigger model for document-level machine translation. In *ACL (2)*, (pp. 382–386). Citeseer.

Bhatia, P., Ji, Y., & Eisenstein, J. (2015). Better document-level sentiment analysis with discourse parsing. In *EMNLP 2015 (under review)*.

Biran, O. & McKeown, K. (2013). Aggregated word pair features for implicit discourse relation disambiguation. In *Proceedings of the Association for Computational Linguistics (ACL)*, (pp. 69–73)., Sophia, Bulgaria.

Burstein, J., Tetreault, J., & Andreyev, S. (2010). Using entity-based features to model coherence in student essays. In *Human language technologies: The 2010 annual conference of the North American chapter of the Association for Computational Linguistics*, (pp. 681–684). Association for Computational Linguistics.

Carlson, L., Marcu, D., & Okurowski, M. (2003). Building a discourse-tagged corpus in the framework of rhetorical structure theory. In J. Kuppevelt & R. Smith (Eds.), *Current and New Directions in Discourse and Dialogue*, volume 22 of *Text, Speech and Language Technology* (pp. 85–112). Springer Netherlands.

Carpuat, M. & Simard, M. (2012). The trouble with smt consistency. In *Proceedings of the Seventh Workshop on Statistical Machine Translation*, (pp. 442–449). Association for Computational Linguistics.

Chardon, B., Benamara, F., Mathieu, Y., Popescu, V., & Asher, N. (2013). Measuring the effect of discourse structure on sentiment analysis. In *Computational Linguistics and Intelligent Text Processing* (pp. 25–37). Springer.

Chen, H., Branavan, S., Barzilay, R., & Karger, D. R. (2009). Content modeling using latent permutations. *Journal of Artificial Intelligence Research*, *36*(1), 129–163.

Ding, C., Utiyama, M., & Sumita, E. (2014). Document-level re-ranking with soft lexical and semantic features for statistical machine translation.

Elsner, M. & Charniak, E. (2010). Disentangling chat. *Computational Linguistics*, *36*(3), 389–409.

References II

Feng, V. W. & Hirst, G. (2012). Text-level Discourse Parsing with Rich Linguistic Features. In *Proceedings of the Association for Computational Linguistics (ACL)*, Jeju, Korea.

Feng, V. W. & Hirst, G. (2014). A linear-time bottom-up discourse parser with constraints and post-editing. In *Proceedings of the Association for Computational Linguistics (ACL)*, (pp. 511–521)., Baltimore, MD.

Ghosh, S., Riccardi, G., & Johansson, R. (2012). Global Features for Shallow Discourse Parsing. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*.

Guzmán, F., Joty, S., Màrquez, L., & Nakov, P. (2014). Using discourse structure improves machine translation evaluation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, (pp. 687–698)., Baltimore, MD.

Halliday, M. & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Hearst, M. A. (1997). Texttiling: Segmenting text into multi-paragraph subtopic passages. *Computational linguistics*, *23*(1), 33–64.

Heilman, M. & Sagae, K. (2015). Fast rhetorical structure theory discourse parsing. arXiv.

Hernault, H., Prendinger, H., duVerle, D. A., & Ishizuka, M. (2010). HILDA: A Discourse Parser Using Support Vector Machine Classification. *Dialogue and Discourse*, *1*(3), 1–33.

Hogenboom, A., Frasincar, F., de Jong, F., & Kaymak, U. (2015). Using rhetorical structure in sentiment analysis. *Communications of the ACM*, *58*(7), 69–77.

Jansen, P., Surdeanu, M., & Clark, P. (2014). Discourse complements lexical semantics for non-factoid answer reranking. In *Proceedings of the Association for Computational Linguistics (ACL)*, Baltimore, MD.

Ji, Y. & Eisenstein, J. (2014). Representation learning for text-level discourse parsing. In *Proceedings of the Association for Computational Linguistics (ACL)*, Baltimore, MD.

Ji, Y. & Eisenstein, J. (2015). One vector is not enough: Entity-augmented distributional semantics for discourse relations. *Transactions of the Association for Computational Linguistics (TACL)*, *3*, 329–344.

Joty, S., Carenini, G., Ng, R., & Mehdad, Y. (2013). Combining Intra- and Multi-sentential Rhetorical Parsing for Document-level Discourse Analysis. In *Proceedings of the Association for Computational Linguistics (ACL)*, Sophia, Bulgaria.

References III

Karamanis, N., Mellish, C., Poesio, M., & Oberlander, J. (2009). Evaluating centering for information ordering using corpora. *Computational Linguistics*, *35*(1), 29–46.

Kibble, R. (1999). Cb or not cb? centering theory applied to nlg. In *Proceedings of ACL workshop on Discourse and Reference Structure*.

Kong, F., Zhou, G., & Zhu, Q. (2009). Employing the centering theory in pronoun resolution from the semantic perspective. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, (pp. 987–996)., Singapore.

Lascarides, A. & Asher, N. (2007). Segmented discourse representation theory: Dynamic semantics with discourse structure. In *Computing meaning* (pp. 87–124). Springer.

Li, J., Li, R., & Hovy, E. (2014). Recursive deep models for discourse parsing. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*.

Li, J. J., Carpuat, M., & Nenkova, A. (2014). Assessing the discourse factors that influence the quality of machine translation. In *Proceedings of the Association for Computational Linguistics (ACL)*, (pp. 283–288)., Baltimore, MD.

Li, S., Wang, L., Cao, Z., & Li, W. (2014). Text-level discourse dependency parsing. In *Proceedings of the Association for Computational Linguistics (ACL)*, Baltimore, MD.

Lin, Z., Ng, H. T., & Kan, M.-Y. (2010). A PDTB-styled End-to-End discourse parser. *arXiv*.

Lin, Z., Ng, H. T., & Kan, M.-Y. (2011). Automatically Evaluating Text Coherence Using Discourse Relations. In *Proceedings of the Association for Computational Linguistics (ACL)*, (pp. 997–1006)., Portland, OR.

Lin, Z., Ng, H. T., & Kan, M. Y. (2014). A PDTB-styled end-to-end discourse parser. *Natural Language Engineering*, *FirstView*, 1–34.

Manabu, O. & Kouji, T. (1996). Zero pronoun resolution in japanese discourse based on centering theory. In *Proceedings of the 16th conference on Computational linguistics-Volume 2*, (pp. 871–876). Association for Computational Linguistics.

Marcu, D. (1999). Discourse trees are good indicators of importance in text. *Advances in automatic text summarization*, 123–136.

References IV

Marcu, D., Carlson, L., & Watanabe, M. (2000). The automatic translation of discourse structures. In *Proceedings of NAACL*.

McKnight, L. & Srinivasan, P. (2003). Categorization of sentence types in medical abstracts. In *AMIA Annual Symposium Proceedings*, volume 2003, (pp. 440). American Medical Informatics Association.

Meyer, T., Hajlaoui, N., & Popescu-Belis, A. (2015). Disambiguating discourse connectives for statistical machine translation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *23*(7), 1184–1197.

Meyer, T., Popescu-Belis, A., Hajlaoui, N., & Gesmundo, A. (2012). Machine translation of labeled discourse connectives. In *Proceedings of the Tenth Biennial Conference of the Association for Machine Translation in the Americas (AMTA)*, number EPFL-CONF-192524.

Oza, U., Prasad, R., Kolachina, S., Sharma, D. M., & Joshi, A. (2009). The hindi discourse relation bank. In *Proceedings of the third linguistic annotation workshop*, (pp. 158–161). Association for Computational Linguistics.

Palau, R. M. & Moens, M.-F. (2009). Argumentation mining: the detection, classification and structure of arguments in text. In *Proceedings of the 12th international conference on artificial intelligence and law*, (pp. 98–107). ACM.

Park, J. & Cardie, C. (2012). Improving Implicit Discourse Relation Recognition Through Feature Set Optimization. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, (pp. 108–112)., Seoul, South Korea. Association for Computational Linguistics.

Pitler, E., Louis, A., & Nenkova, A. (2009). Automatic Sense Prediction for Implicit Discourse Relations in Text. In *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP* .

Pitler, E. & Nenkova, A. (2008). Revisiting readability: A unified framework for predicting text quality. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, (pp. 186–195). Association for Computational Linguistics.

Pitler, E., Raghupathy, M., Mehta, H., Nenkova, A., Lee, A., & Joshi, A. (2008). Easily identifiable discourse relations. In *Coling 2008: Companion volume: Posters*, (pp. 87–90)., Manchester, UK. Coling 2008 Organizing Committee.

Poesio, M., Stevenson, R., Di Eugenio, B., & Hitzeman, J. (2004). Centering: A parametric theory and its instantiations. *Computational linguistics*, *30*(3), 309–363.

References V

Poláková, L., Mírovský, J., Nedoluzhko, A., Jínová, P., Zikánová, v., & Hajičová, E. (2013). Introducing the prague discourse treebank 1.0. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, (pp. 91–99)., Nagoya, Japan. Asian Federation of Natural Language Processing.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., & Webber, B. (2008). The penn discourse treebank 2.0. In *Proceedings of LREC*.

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

Prasad, R., McRoy, S., Frid, N., Joshi, A., & Yu, H. (2011). The biomedical discourse relation bank. *BMC bioinformatics*, *12*(1), 188.

Prasad, R., Webber, B., & Joshi, A. (2014). Reflections on the penn discourse treebank, comparable corpora, and complementary annotation. *Computational Linguistics*, *40*(4), 921–950.

Rutherford, A. T. & Xue, N. (2014). Discovering Implicit Discourse Relations Through Brown Cluster Pair Representation and Coreference Patterns. In *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL)*, Stroudsburg, Pennsylvania. Association for Computational Linguistics.

Sagae, K. (2009). Analysis of Discourse Structure with Syntactic Dependencies and Data-Driven Shift-Reduce Parsing. In *Proceedings of the 11th International Conference on Parsing Technologies (IWPT'09)*, (pp. 81–84)., Paris, France. Association for Computational Linguistics.

Somasundaran, S., Wiebe, J., & Ruppenhofer, J. (2008). Discourse level opinion interpretation. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, (pp. 801–808). Association for Computational Linguistics.

Soricut, R. & Marcu, D. (2003). Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of NAACL*.

Sporleder, C. & Lapata, M. (2005). Discourse chunking and its application to sentence compression. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, (pp. 257–264). Association for Computational Linguistics.

Teufel, S. (1999). *Argumentative zoning: Information extraction from scientific text*. PhD thesis, University of Edinburgh.

References VI

Tu, M., Zhou, Y., & Zong, C. (2013). A novel translation framework based on rhetorical structure theory. In *Proceedings of the Association for Computational Linguistics (ACL)*, (pp. 370–374)., Sofia, Bulgaria. Association for Computational Linguistics.

Tu, M., Zhou, Y., & Zong, C. (2014). Enhancing grammatical cohesion: Generating transitional expressions for smt. In *Proceedings of the Association for Computational Linguistics (ACL)*, (pp. 850–860)., Baltimore, Maryland. Association for Computational Linguistics.

Wang, W., Su, J., & Tan, C. L. (2010). Kernel Based Discourse Relation Recognition with Temporal Ordering Information. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*.

Webber, B. (2004). D-LTAG: extending lexicalized TAG to discourse. *Cognitive Science*, *28*(5), 751–779.

Wolf, F. & Gibson, E. (2005). Representing discourse coherence: A corpus-based study. *Computational Linguistics*, *31*(2), 249–287.

Wong, B. & Kit, C. (2012). Extending machine translation evaluation metrics with lexical cohesion to document level. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, (pp. 1060–1068). Association for Computational Linguistics.

Xuan Bach, N., Le Minh, N., & Shimazu, A. (2012). A reranking model for discourse segmentation using subtree features. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, (pp. 160–168)., Seoul, South Korea. Association for Computational Linguistics.

Zeyrek, D., Demirşahin, I., Sevdik-Çalli, A., Balaban, H. Ö., Yalçinkaya, İ., & Turan, Ü. D. (2010). The annotation scheme of the turkish discourse bank and an evaluation of inconsistent annotations. In *Proceedings of the fourth linguistic annotation workshop*, (pp. 282–289). Association for Computational Linguistics.

Zhou, Y. & Xue, N. (2012). Pdtb-style discourse annotation of chinese text. In *Proceedings of the Association for Computational Linguistics (ACL)*, (pp. 69–77)., Jeju, Korea.